

## NRC Publications Archive Archives des publications du CNRC

### The social context of artificial intelligence: a guideline and discussion paper: abridged version National Research Council Canada

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

<https://doi.org/10.4224/23001517>

**NRC Publications Archive Record / Notice des Archives des publications du CNRC :**  
<https://nrc-publications.canada.ca/eng/view/object/?id=cd35d0cc-59b4-49c3-beeb-787763828019>  
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=cd35d0cc-59b4-49c3-beeb-787763828019>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at <https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site <https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at [PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca](mailto:PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca). If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à [PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca](mailto:PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca).

Q335  
S678  
c. 2



National Research  
Council Canada

Associate Committee  
on Artificial Intelligence

Conseil national  
de recherches Canada

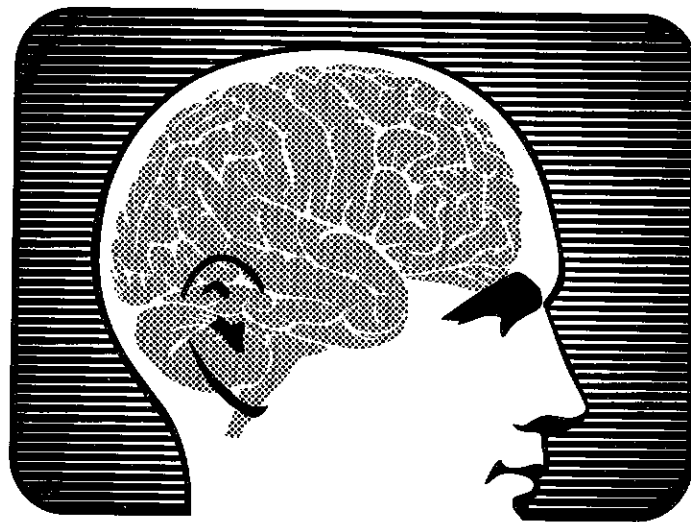
Comité associé de  
l'intelligence artificielle

---

# ***The Social Context of Artificial Intelligence***

A Guideline and Discussion Paper  
Abridged Version

June 1989



Copyright 1989 by  
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

This report, NRCC No. 30188, is a condensed version of the full report NRCC No. 30187 which bears the same title. Additional copies of both are available free of charge from:

Editorial Office, Room 301  
Division of Electrical Engineering  
National Research Council of Canada  
Ottawa, Ontario, Canada  
K1A 0R6

Copyright 1989 par  
Conseil national de recherches du Canada

Il est permis de citer de courts extraits et de reproduire des figures ou tableaux du présent rapport, à condition d'en identifier clairement la source.

Ce rapport, CNRC n° 30188, est une version abrégée du rapport complet CNRC n° 30187 du même titre. Des exemplaires supplémentaires gratuits des deux rapports peuvent être obtenus à l'adresse suivante :

Bureau des publications, Pièce 301  
Division de génie électrique  
Conseil national de recherches du Canada  
Ottawa (Ontario) Canada  
K1A 0R6

Disponible en français, CNRC n° 30463.

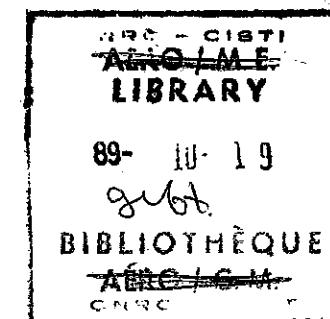
---

# *The Social Context of Artificial Intelligence*

A Guideline and Discussion Paper:  
Abridged Version

National Research Council of Canada  
Associate Committee on Artificial Intelligence

June 1989



Canada Institute for Scientific  
and Technical Information  
National Research Council  
Canada  
Institut canadien de  
l'information scientifique et  
technique  
Conseil national de recherches  
Canada

\*  
Q335  
S678  
C.2  
~~ABD~~



## Table of Contents

Preface .....	v
I — Introduction .....	1
II — Social Considerations Arising from Artificial Intelligence .....	9
Introduction .....	11
Issues and Concerns Arising from Artificial Intelligence .....	11
Work .....	12
Decision Making .....	12
Social Organization .....	13
Situation Examples and Discussions .....	13
A Difference between AI and Data Processing Applications .....	16
Regulatory Bodies .....	17
An Industry Association for Voluntary Accreditation of AI Products ..	17
Codes of Ethics .....	18
Validation .....	18
Traceability .....	19
Documentation .....	19
Summary and Conclusions .....	19
III — A Protocol for the Examination of AI Applications from the Social Point of View .....	21
Contexts for a Case Study .....	23
The Focal Organizational Context .....	23
IV — Summary and Recommendations .....	27
Summary of the Protocol .....	29
Additional Issues, Concerns, and Questions to Ask .....	30
Summary of Issues and Concerns .....	32
Conclusions .....	33
V — References .....	35
Appendix 1 — Terms of Reference: NRC Associate Committee on Artificial Intelligence .....	39
Subcommittee on the Social Context of Artificial Intelligence .....	41

## **Preface**

Associate Committees established by the National Research Council Canada have been used for over fifty years on many topics to act as a focal point and forum in Canada for issues of national concern to the engineering and scientific communities. The NRC Associate Committee on Artificial Intelligence (ACAI) was established in April 1987. In its first meetings the ACAI established a sub-committee on the social context of AI, recognizing that artificial intelligence carries with it, not only important technical and economic considerations, but also considerations of a social nature.

Examination of AI in the social context reveals that there are considerations and choices of a social nature related to the way AI may be applied in the workplace, there are social considerations related to the professional responsibility for AI applications and there are large scale macro effects with the potential to affect or influence society as a whole.

The subject of artificial intelligence is bounded on one side by the general subject of information technologies (of which it is a part) and on the other side by the general subject of industrial automation. A very large volume of literature exists on the social aspects of information technologies in general. Similarly, much is available on the social implications of industrial automation. Both of these fields have much longer histories than AI, and it is useful to draw some examples or to point to some of the references available from them. To maintain a forward focus, however, an attempt has been made in this report to restrict most of the considerations presented to the field of artificial intelligence itself.

Near the end of the first meeting of those directly involved in the preparation of this document, one contributor remarked on the limitations of such an endeavour.

"I don't know" he said, "if a group of wisemen ever gathered around the steam engine when it

was first invented and if they sought to predict all the social effects that would follow". History has shown us how nearly impossible such an attempt would have been. Nevertheless, in the present instance of artificial intelligence, the contributors to this report believe it is worth something to try.

By producing this report, which includes a protocol for the examination of AI applications from the social point-of-view, it is the hope of those who have contributed to it that our understanding of artificial intelligence will be enhanced, and that the selection and design of AI applications will be undertaken with an improved understanding of the social considerations attached to them in addition to those considerations of a business and technical nature.

Understandably, neither this report nor its companion report NRCC No. 30187 are intended to provide answers and solutions to the many questions asked or suggested by the situations described. They may provide some degree of guidance, however, by seeking to anticipate them, and presenting them in a balanced and structured manner. The answers or solutions eventually will come from many sources and will only evolve with time and experience. Even the real questions will only evolve in a similar manner.

Since this is a first report, correspondence is invited and will be acknowledged. While it may not be possible to enter into discussion on all communications received, they will be taken into account in any future report or addendum.

Correspondence should be addressed to:

Artificial Intelligence in the Social Context  
NRC Associate Committee on Artificial Intelligence  
Room 205, Building M-50  
National Research Council of Canada  
Ottawa, Ontario  
K1A 0R6

*Section I*

---

*Introduction*

## Section I — Introduction

Before one can begin to consider the business, technical or social ramifications of artificial intelligence, it is necessary to have some reasonable definition of what is meant by the term "artificial intelligence".

Artificial intelligence (AI) is not easily defined. It may be considered to be founded on a branch of computer science that concerns the conception, design, development, and application of computer-based systems to perform tasks that require the exercise of intelligence as exemplified by capacity to recognize, reason, decide, and learn. Thus application of AI seeks to incorporate such human-like skills and expertise into computer-based systems. These systems include what are called knowledge-based and expert systems.

It is an essential characteristic of an application of AI that the associated computer system has provision for representing definable elements of knowledge about the context in which the system will be used and about the goals intended to be realized in its use. Further, when the range of conditions of application of the system is appropriately limited, the behaviour exhibited by the system is at least intended to be rationally attributable to the content within it of the representation of its knowledge and goals. These characteristics of AI systems will be illustrated by a variety of examples.

Computer systems which realize artificial intelligence commonly display advanced forms of input and output such as:

- Speech input and voice recognition
- Natural language processing
- Machine vision
- Photo interpretation
- Image understanding
- Advanced forms of sensor-based input
- Voice synthesis and speech output
- Mechanical and manipulation output capability (as in robotics).

Applications of AI include, but are not limited to, the following:

- Natural language processing
- Machine translation

- Expert systems, such as those used in the diagnosis and treatment of disease, monitoring and diagnosing faults in equipment, industrial plants, telecommunications, and electric utility systems
- Operations scheduling and process planning in manufacturing
- Robot and other autonomous systems with vision or other sensing capabilities that permit them to respond to changes in their external environment.

Expert systems are of particular interest because in terms of application they represent a leading edge of AI. They usually exhibit most, but not necessarily all, of the attributes listed in Table I.

A short list of potential application areas is shown in Table II in order to illustrate the breadth and pervasive nature of artificial intelligence. It is necessary to recognize this breadth in order to place the social context of AI in perspective. If the applications of AI become as widespread as some advocates believe, or if even only partially so, its influence will be significant on the daily lives of a large number of people. A purpose of this report is to show in what ways this may happen, and to suggest how the benefits may be maximized while minimizing any undesirable effects.

For example, a set of technical and economic factors that may be used in the selection of expert system applications is shown in Table III. By developing a parallel set of factors for the examination of AI applications from the social point of view (such as those contained in Section III of this report) an additional set of useful criteria may be brought into play.

This report, in order to meet its purpose of facilitating and enhancing the process of adaptation, must necessarily concentrate or focus on situations where system and technological changes will be needed in order to meet human and social needs, or on situations where social adaptation and organizational change may be needed as adjustments to the new technological capability. While the report must necessarily focus on these issues, it is necessary at the same time to retain a clear image of the benefits which will be obtained from the continued development and application of artificial intelligence. These will be

Table I. Expert System Attributes

**Nature of Application and Use**

Expert systems are used to provide advice in much the same way that it would be obtained and used if a human expert were available.

**Domain expert**

In the building of an expert system it is necessary to have available, and to draw on, the knowledge and experience of a person who is already an expert in the subject area. This person is referred to as the domain expert.

**Rule based**

Rather than being based on procedures written in algorithmic form, many expert systems derive their knowledge from a collection of rules.

**Method of knowledge representation**

The system knowledge is contained and given by the rules in the knowledge base rather than being represented by procedures given by computer coding as in "conventional" programs.

**Symbolic representation**

Expert systems tend to work with information in symbolic form rather than being based on calculations performed in numeric form. Use may be made of input, output or rules expressed in a natural language such as English, rather than being expressed numerically.

**Search strategies**

In seeking a problem solution, expert systems examine, select, and manipulate rules in the knowledge base by means of search strategies such as forward and backward chaining. The portion of the system performing the logic function in an expert system is sometimes known as the inference engine.

**Computer language**

Expert systems are often written in list processing languages such as LISP, or logic programming languages such as Prolog, rather than in procedurally oriented languages such as Fortran.

**Traceability**

An important attribute of expert systems is their ability to explain how a given result was obtained. Also known as transparency or explanation facility.

**Apprenticeship**

There is often a need to subject expert systems to a period of apprenticeship as part of their development because it is harder to verify and validate expert systems than it is for computer programs of a numerical or computational nature.

**Designed to grow incrementally**

Since the knowledge base is separate from the search logic, it should be possible to add new rules or modify existing rules without having to review or re-organize the original structure.

Table II. Application Areas for AI and Expert Systems (1)

**Manufacturing**

- Design Synthesis and Optimization
- Process Planning
- Process Parameter Selection, e.g., Robot Arc Welding
- Robotic Vision and Intelligent Robot Systems

**Process Control**

- Analysis and Reduction of Alarm Conditions
- Defining Process Management Rules
- Modelling the Performance of Best Operators

**Materials**

- Selection of Best Materials for a Given Design Function
- Selection of Coating Materials and Finishes

**Resource Industries**

- Evaluation of Exploration Data
- Crop and Herd Management in Agriculture
- Mineral Prospecting
- Mine Planning and Design
- Operations Management

**Maintenance**

- Diagnosis of Faults in Plant and Office Equipment

**Health Care**

- Medical Diagnosis

- Determining Optimum Treatment Cycles, e.g., as in Radiation for Cancer

**Education**

- Intelligent Tutoring
- AI-based Authoring Environments
- Discovery Environments
- Educational Interfaces

**Career Guidance**

- Advising on Education Paths needed for Employment Desired in Specified Occupations

**Transportation**

- Air Traffic Control
- Equipment Maintenance
- Inventory Control

**Financial Planning**

- Advice on Investment Decisions and Planning

**Space**

- Remote Sensing and Data Interpretation

**Environment**

- General Weather Forecasting
- Prediction of Hail Storms

**Automated Language Translation****Speech Recognition****Natural Language Processing****Military Battle Management**

Table III. Criteria for Choosing an Expert System Application Area (2)

- Good Payback
- Well-Bounded Problem
- Real Problem
- Small, Manageable Problem or Application Area
- Stable Proven Technology Available
- Expertise Available

considerable, widespread, and varied, because that is inherent in the nature of AI.

Basically the benefits of AI tend to be derived in the following three ways:

1. Knowledge-based systems tend to capture, represent, and distribute knowledge more uniformly throughout organizations and societal structures. Industrial firms, for example, find that the knowledge and experience gained by long term employees can be retained and distributed throughout the organization in a more uniform and accessible way than was previously possible.
2. Improved decision making is achieved via the quality of the knowledge base and reasoning process employed, and also in the speed of response. Systems are already being developed for a wide variety of applications, particularly in the form of expert systems.
3. As a result of improved and more natural interfaces and means of communication between people, between systems, and between people and computers. The possibility of natural language processing, the possibility of improved methods of machine translation for the many written and spoken languages of the world, the possibilities envisaged from computer recognition of speech and machine vision are all examples of this nature.

Although definitions exist for what is meant by "artificial intelligence", examples constitute one of the best ways by which to gain an understanding of it. The full report, of which this is a condensation and overview, therefore presents brief summaries of papers published in the field of AI. These describe many of the concepts, current developments, and application areas of AI, and illustrate its widespread and pervasive nature. These examples, selected from over 700 recent AI publications, and with emphasis on the practical side, explain much of what AI people plan, propose, and do.

At the same time these illustrate the many sources from which the benefits of AI will be derived, by making the knowledge and experience of experts more widely available, as in expert systems, and by making computers more accessible, as in the many forms of new input/output interface.

By way of illustration, the references abstracted and summarized in the full report include the following topics:

#### *Theoretical Concepts*

- By studying the mechanisms of the brain, research on neural nets may lead to self-learning systems.
- The influence of language on the ability to reason correctly.
- A comparison of six leading methods for reasoning in the presence of uncertainty and imperfect knowledge.
- Methods for knowledge acquisition by inductive example.
- Why attempts to formalize common sense knowledge have failed.
- A code of conduct for professionalism in AI.
- A probabilistic system for inductive learning and predicting future events.

#### *Systems and Techniques*

- The use of AI to extend and enhance conventional information systems.
- How to handle the problem of uncertainty in medical diagnostic systems.
- Variations incurred in expert systems may often be due to the choice of domain expert.
- Establishing company/university joint ventures in AI.
- How to select the correct expert system shell for a given application.
- Human and organizational problems related to expert system applications.
- Difficulties with the maintenance of the knowledge base in an expert system.
- A comparison of how expert system shells deal with uncertainty.
- A practical guide to aid knowledge engineers in the process of interviewing domain experts.

- An expert system for the verification of nuclear test ban treaties by the interpretation of seismic data.
- Techniques for improving the knowledge acquisition process by means of induction.

#### *Applications*

- An expert system for advice in the handling of hazardous waste.
- A retrieval system for information about patents.
- A hierarchical expert system for the synthesis of chemical process flowsheets.
- Extensions, by means of expert systems, to the preparation of NC machine tool programs.
- Monitoring the health of rocket engines by integrating vibration analysis and pattern recognition techniques with AI.
- Air traffic control using AI techniques.
- Legal and contractual ramifications of expert systems.
- A system for evaluating project proposals to a government funding program.
- Military and defence applications of expert systems.
- A financial investment assistant for portfolio management.
- A system for screening applications for university admission.
- Providing advice to new and small businesses.
- A system to advise students in career planning.
- Mineral identification by the analysis of the infrared reflectance spectra of rock samples.
- A system for diagnosing faults in telecommunications switching systems.
- A system to assist engineers in setting up problems for structural analysis using the nonlinear finite element method.
- An example of neural network training; driving a vehicle through simulated freeway traffic.
- Expert system applications in the NASA space station.
- The use of expert systems in architecture, engineering, and construction (AEC).

*Section II*

---

*Social Considerations Arising from  
Artificial Intelligence*

## ***Section II — Social Considerations Arising from Artificial Intelligence***

### **Introduction**

There are both micro and macro effects associated with the development and use of almost any technology. Micro effects are those that directly affect the individual worker or user. Macro effects usually evolve more slowly over time and tend to affect large numbers of people indirectly. This often occurs through changes in the societal and economic structure brought about by the technology in question.

For example, in the case of office automation systems, the widespread use of video display units (VDUs) can have a direct effect on the comfort or physical and mental strain of workers who use them for long periods of time, particularly if sufficient attention is not given to ergonomic factors such as lighting, and work station layout in the systems design.

At an intermediate level the use of credit cards, made possible by computer technology, provides a useful illustration. Not only have credit cards obtained widespread acceptance as a convenience to the user, but also they have fostered the establishment of new credit and service companies. These have altered the structure of the financial community. Additionally, the spending habits of people have changed (many in a substantial way), in addition to the simple and direct convenience of the card itself as a means of payment. These are macro effects, brought about by credit cards and computer technology.

As an example of a very wide scale macro effect, some writers see the possibility that the widespread use of computer systems will gradually alter democracy as we know it today. This is predicted because of a slight but perceived bias in systems design which favours a slow drift towards centralization of power rather than favouring the individual.

The field of industrial automation and computer integrated manufacturing offers another, and perhaps clearer, example of both micro and macro effects. Workers such as welders in the automotive industry have already experienced job displacement, although not necessarily widespread job loss, as a direct effect of robot welding applications. On the other hand, as a macro effect, the

plant that makes no technological change, and which suddenly finds itself to be uncompetitive, poses an even greater employment risk. All workers in the company could be adversely affected. Even the standard of living of entire nations can be impacted if their industries somehow fail to keep up with technology and thereby lose competitive position.

In this section, the report will develop and illustrate more fully the subject of social issues, at both the micro and macro level, related to artificial intelligence applications. To place AI in perspective, and to draw from more traditional information technology applications, a section of the full report provides a summary view of social considerations which arise from information technologies other than artificial intelligence. These are closely related, and indeed often impossible to separate from those of AI itself.

### **Issues and Concerns Arising from Artificial Intelligence**

It is difficult to distinguish between the social implications stemming from the use of computers in general, and those arising out of artificial intelligence applications in particular. Most AI applications are still experimental and the social implications will only become clearer when there is a more general use. If its potential is realized, AI may usher in a new age which will truly be a qualitative break with the past. Only the future will tell, but in any event early recognition of the issues involved can serve to ease the transition.

Clearly there are many applications of intelligent systems of direct and unequivocal benefit to society at large and for workers in particular. For example dangerous activities in mines, under the sea, in nuclear plants, in the chemical industry, and elsewhere are prime candidates for robots. Other jobs, less dangerous, but obviously unpleasant, should in the near future be also phased out as exclusively human preserves.

AI is a vital and exciting field that has attracted many scientists. Research at universities, in private laboratories and in government installations is proceeding across many fronts. These include knowledge representation, reasoning, problem solving, natural language understanding, image processing, expert systems, logic programming, and heuristic search, among others.

It can be argued that the widespread use of expert systems brings to the fore some very serious issues. By their very nature, expert systems purport to capture, formalize, and disseminate expertise. From a societal point of view, the effects of this process may include standardization, homogenization, centralization, legitimization, and a definite sense of authority and control. Part of the concern expressed in some instances is that the formalization of knowledge as an expert system for some restricted domain can be taken as the representation of the knowledge; on the other hand in many areas of life there is no consensus, no received view and, in the opinion of many, there will never be. Thus a threat possibly posed by the rapid diffusion of expert systems is a limitation of diversity and the imposition of the equivalent of a state religion, with its fixed, dogmatic worldview.

While an advantage of expert systems lies in their ability to disseminate knowledge and make the ability of experts more widely available, difficulties with updating and maintaining the associated knowledge bases is certain in some instances to create a state of frozen and inflexible logic. This could exacerbate the difficulty presented above. Four specific issues may be chosen to illustrate some of the social considerations arising from artificial intelligence technology (3). These involve the influence or inter-relationship between AI and:

- Work
- Decision making
- Privacy
- Societal structure, ethics, and political adaptation

### Work

Of particular concern is the impact of computers on work — both the nature of the job itself and the number of jobs. The relation between technology and work is complicated and operates on many dimensions. The economic imperative to introduce new technology and to increase productivity in order to be competitive is alive and well today as it has been since the onset of the Industrial Revolution and before.

However, the emergence of AI will bring to the fore the question: Will there be a massive loss of jobs and if so, what kinds of jobs will be available? Various writers, including more recently AI

researchers, have speculated about a future in which intelligent machines produce the goods and provide many of our services. Without pursuing this topic further, it is noted that serious issues related to income distribution, self-worth, and the basic political organization of society may be involved.

Surely intelligent machines can be usefully employed to perform tasks that are undesirable for people, for example, both underground and under-sea mining, dealing with hazardous wastes, welding and spray painting, and handling noxious gases. Of greater significance is the intellectual benefit of intelligent systems for improving efficiency in every aspect of human endeavour. In general, knowledge-based systems can be used for applications in the executive suite and on the factory floor. Intelligent aids to information retrieval, decision making, planning, and problem solving are appearing in almost every conceivable area of application. One can only expect that they will continue to improve and be even more widely applied.

### Decision Making

This term is used here in an all-encompassing sense to cover activities regularly carried out by individuals, companies, institutions, and governments. Every aspect of life involves decisions, whether made by the individual or made for him, or her, by others. As such, decision making represents a fundamental component of human existence and threats to human autonomy, however couched in friendly terms, are of serious concern. The question of the dogmatic aspect of formalized expertise contained within expert systems has already been raised. Indeed, the increasing use of computers must inevitably result in a decrease in individual human decision making. The role of AI will be to accelerate this trend, especially in more critical situations. Of special mention in this respect, the debate over the Strategic Defense Initiative (SDI, or more commonly, Star Wars) has frequently hinged on the question of whether or not the very large software component could perform as required. This system will monitor information-gathering devices, assimilate the information, decide on a response, coordinate the response, and continue with these activities until the end. AI is certainly expected to play a role in this system. Note that the issue of computer decision making does not begin with AI but rather

that the reliance on AI may exacerbate the potential problem in a fundamental way.

There are a number of benefits associated with expert systems and these are important in their own right, not merely to offset the possible dangers outlined above. Expertise is a rare commodity and the availability of mechanisms to preserve and disseminate it are to be valued. Expert systems can serve as a means of preserving knowledge, in an active and useful way, even for future generations. They can ensure that the knowledge gained over a lifetime by an expert can be saved in a useful form. Expert systems can also perform an educational service, beyond traditional text books, in educating the next generation of experts.

Expert systems may yield substantial benefits in such areas as medical emergencies, poison control, and other life-threatening situations. Rapid access to the specialized knowledge contained in expert systems may make the difference between life and death in these situations. In remote locations such knowledge may not be readily available from people, from books, or even from computer databases. Appropriate expert systems can substitute for human specialists under some circumstances. Who could argue against their use in emergency conditions, but concern may arise, however, when expert systems diffuse into more mundane or less essential areas.

### Social Organization

How will society, or better its political institutions, respond to a future in which basic needs of both goods and services are met by machines? Robots and advanced industrial automation are gradually reducing the blue collar workforce. This is taking place largely without AI. Changes are occurring much more slowly within the office but the impact of successful developments in AI, especially in speech understanding, could result in many fewer jobs. Thus a key issue for the future may be what will replace work in most people's lives both as a means to acquire wealth and as a major component in the definition of self-worth.

Since the most widely accepted way to distribute real wealth in society is through wages or salary, and since the envisioned future includes a considerably reduced workforce, two questions emerge: How will people acquire the means necessary to acquire goods and services beyond immediate

basic needs and what will replace work as the major activity in most people's lives?

It is not obvious how the social and economic adaptation to such a future will evolve or even that it will. It appears that massive changes in political systems as well as in social organization will be necessary. Work and money are just part of the equation. Autonomy, self-respect, and civil liberties are others. None of these are gifts bestowed by a benevolent state, especially one which is the product of major technological innovations. It is for this reason that the process involved in moving towards a new society is so crucial and that an awareness and realistic understanding of how technology operates, perhaps the philosophy of technology itself, is so important.

At the same time, if AI is seen as a natural and expected continuation of the historical evolution of technology, then there is no reason to expect its effect to be substantially discontinuous with the past. Some two hundred years have passed since the onset of the Industrial Revolution, yet we still find that unemployment rates are relatively low.

The nature of AI, however, suggests that as a new technology, its influence may not be merely quantitative, not just more of the same, not just the latest improvement in the continuous progress of the industrial revolution. If its potential is realized, AI may usher in a distinctly new age, and ultimately it will truly be a qualitative break with the past. The question then becomes one concerned with the rate of change, and the time available for adaptation.

### Situation Examples and Discussions

Four examples have been chosen from many that are possible, to illustrate more clearly and in greater detail some of the social considerations related to the future use of expert systems.

The first example, in realtime process control, if not describing what is already a real situation, describes one that may soon be real. The second example, in the field of medical diagnostics, further develops a number of issues of concern and places them in the realm of direct personal considerations. A third example indicates an instability that may exist if programmed trading in stock market securities reaches the level of widespread practise. A fourth example is the possible

initiation of a nuclear war due to an error in computer decision making.

#### **Situation 1 — Realtime Process Control**

Assume that a realtime expert system is installed to aid the operator in the control room of a large refinery or petrochemical plant in assigning priority to alarm signals during a major emergency or process upset. Without the expert system, the operator is certain, at these times, to be flooded with alarm signals and notifications of "out-of-limits" conditions, some of which are much more important than others. With the expert system it is hoped that the operator will be in a better position to identify the real cause of the breakdown, and to more quickly determine the corrective action to be taken. The problem, however, is that everything happens fast, in realtime, and the operator may have little choice other than to follow the advice of the expert system.

Who is responsible then if a major upset occurs and the operator follows the expert system's advice but it is erroneous because:

- the expert system rules and analysis were incomplete
- the process upset (or combination thereof) was something never considered and therefore outside the domain of the expert system
- it had actually been developed for a different kind of plant
- there was a logic error in the software
- there was an interface error to parts of the system developed by others
- some sensors had been damaged by the accident and therefore the expert system could not function
- the expert system functioned, but assumed that input data and signals were always complete and correct when in fact they were incomplete or erroneous
- etc.

One could argue that the example provided is not significantly different from traditional control systems with which there is already a wide and generally satisfactory experience, or that it does not differ from expert systems in general. The

example is chosen, however, because it does differ in two important ways:

- Expert systems are less deterministic in their performance than sequential computer programs. It is more difficult to predict or test their behaviour and output over the range of input conditions. By their very nature, many expert systems base their results on inferences rather than certainties. Normally the operator may be able to make better judgements and to perform better with the expert system than without it, but there may be no guarantee of this under all conditions.
- In the example given above the expert system operates in realtime and is expected to be used in situations where the human operator has little or no opportunity to cross-check or evaluate its output. He or she will be very dependant on it, and could be forced by circumstances to trust it implicitly. Who, then, bears responsibility to the public for the plant's operation?

This example, from a perhaps hypothetical industrial environment, is sufficient to illustrate the considerations involved. There are parallel considerations for many potential military applications of expert systems or for any artificial intelligence applications which would be expected to operate in realtime, in a variety of situations, where complete pre-testing would be very difficult to arrange.

#### **Situation 2 — Medical Diagnostics**

A considerable number of the early and best known expert systems are in the field of medical diagnosis. The reasons for their development can be several fold:

- as aids to teaching the methodology of medical diagnosis,
- to collect and pool expert knowledge in areas where diagnosis is particularly difficult,
- to study a typical application area in order to further develop problem understanding and the AI techniques required for its solution,
- possibly for actual application to diagnosis in clinical situations.

It is in the latter area of real application where potential problems of a social nature may be

encountered. To clarify what is meant, and to invoke discussion, a number of different, but possible, scenarios are presented below. These are drawn from potential situations in medical practise, but parallel situations could be found in other expert system application areas, particularly within the professions, such as the provision of legal advice, engineering design, and career guidance to name just a few.

Today many expert systems do not necessarily perform well but exist primarily to illustrate their potential, pending further development. However, just as chess playing programs once performed only at the novice level these now perform, sooner than was expected, at the masters level. We need to look forward already to what may happen if, or when, expert systems begin to equal and then surpass the ability and judgement of those we deem to be experts, and whom we trust, today. We may consider therefore, the following:

- A physician has an expert system available to him to aid in the diagnosis of a patient. At least six possibilities immediately arise:
  1. The expert system agrees with his own analysis. The physician proceeds to follow the appropriate treatment procedure with added confidence that it is correct.
  2. The expert system does not agree with his analysis. He proceeds to follow his own analysis, which turns out to be right.
  3. Same as (2), but it turns out that the expert system was right and the physician was wrong.
  4. The expert system does not agree with the physician's analysis. He or she decides to follow the analyses of the expert system, which turns out to be right.
  5. Same as (4), but the expert system is wrong.
  6. Although the expert system was available, the physician chose not to use it. Several situations arise:
    - This was a good choice, human skill and judgement prevailed and no known problems arose.
    - An error in diagnosis did occur and it was discovered that had the expert system been

used and relied upon, the error would not have occurred.

The implications of this latter possibility are even more pronounced if it is considered that the expert system available, by virtue of continued development, may have reached the status of accurately modelling the diagnostic capability of a highly qualified peer group of physicians, exceeding the knowledge and ability of even any single member of that group.

- Yet another series of situations exists if one envisages a situation in a geographically remote region, or in a lesser developed country, where trained physicians are not widely available. The expert system might be marketed and used, even to advantage in the majority of cases, by paramedical personnel, by the patients themselves, or by well-meaning friends acting on their behalf.

Candidates for the assignment of responsibility, especially in instances where errors and problems might have occurred, include:

- the individual physician
- the hospital, clinic or group practise
- the supplier of the expert system (possibly from another country)
- the distributor of the expert system
- the manufacturer of the computer
- the research team that provided the domain expertise and knowledge base
- the developers or supplier of the expert system shell
- the AI team that developed it into a working system for medical diagnosis
- the patient who attempted to use it by his or herself
- the well-intentioned friends of the patient.

It is distinctly possible that in a number of instances the rate of development and application of expert systems, may be limited not by the technology itself but by matters related to legal responsibility. This may be particularly so in the professions. For this reason, some potential applications may only achieve use as teaching aids, rather than being implemented for use in direct practise.

Alternatively, as the technology develops and moves to direct application, the education requirements for professionals deemed responsible for

the use of expert systems may rise to a very high and even extraordinary level. This could severely strain the ability of suppliers to provide appropriate program documentation and training. In fact, it may turn out in some instances that practitioners who genuinely understand the program, and who have the ability to use it correctly, are only available at very special centres of expertise.

These centres may often be located outside any particular country, leading in some instances to a loss of national control. A form of "software colonialism" may develop as funds and resources continue to flow through the marketing chain to these sources of expertise and thereby reinforcing centralized development.

On the other hand, this may not necessarily happen. Just as the world once looked almost solely to Versailles for art, or Hollywood for cinema, other developments over time have caused the production capacity for these two examples of expertise and achievement to become widely distributed. Even though concentration of knowledge may appear to be an initial consequence of expert systems, it may not necessarily continue that way.

#### **Situation 3 — Programmed Trading in the Stock Market**

Programmed trading in the stock market is another area where it has already been recognized that regulatory controls may be required. While individual usage is not deemed to be an area for concern at present, multiple simultaneous use creates the conditions for an avalanche effect with rapid, widespread effects of great consequence. Stock markets, such as the New York Stock Exchange, have already given serious consideration to imposing limits on their immediate members for the use of programmed trading. If the technique and practise should become more widespread throughout the financial community at large than it presumably is at present, an even more unstable condition may result than was experienced in October 1987. Fears of a worldwide economic collapse and recession triggered by such action, while it has not happened, are not totally unrealistic.

Many types of systems, both those created by man and by nature, can demonstrate and even suffer from dynamic instabilities. However, except for electric power networks and feedback control systems, few of these, due to the intractability of

the problem, have formal and well-developed methodologies for stability analysis.

#### **Situation 4 — The Initiation of Nuclear War**

Many computer scientists are concerned with the possibility that a computer error could initiate a nuclear war. The organization known as "Computer Professionals for Social Responsibility" (CPSR) is particularly concerned with this possibility. While many of these concerns stem from questions related to reliability of software in general, particularly the massive software system required in the proposed Strategic Defence Initiative (SDI), some of these concerns could originate from the proposed role of artificial intelligence. An example, in pattern recognition, is the evaluation of sensor data to determine target identification and the overall decision logic used to determine if an apparent event does in fact represent a hostile attack.

#### **A Difference between AI and Data Processing Applications**

In general most of the social issues related to the general or conventional use of computers relate also to AI. Thus it is difficult to separate the two. In some instances the role of AI may be to magnify the effect or concern in question, particularly those related to the involvement of computers in decision making as distinct from their use for numerical calculation.

A few examples will illustrate how important it is to ask the right questions in the examination of artificial intelligence applications. An expert system is described in the literature for allocating children to foster homes. This raises the question, "How can the expert system perform this function better than human judgement?", whereas the correct question to ask is more likely, "Does the human-machine combination perform better in this instance than would be the case if decisions were based on human judgement alone?"

Another point to be recognized in the examination of AI applications is that the correct question to be asked in many instances is not just, "Does the human-machine system perform better, on average, than would be achieved with human judgement alone?", but rather "Even in extreme instances does the human-machine system perform better?" In some cases this means that performance must be almost infallible.

For example, in 1988, in a case of mistaken identity, a commercial airliner was shot down and destroyed by missile fire from the naval vessel of another country. The incident occurred "over the horizon", which means that electronic means of detection were clearly involved. At one point in the enquiries that followed this event it was reported that the mistaken identity may have been caused by an error in software employed for pattern recognition and target identification. Later reports indicate that this was not the case, and that the mistake was due to human operator error (4). Had it been due to the automated pattern recognition system, however, this would have fallen into the domain of artificial intelligence, at least according to some definitions.

This illustrates the need, in many AI applications, for extremely careful system validation over a wide range of operational conditions, such that extreme values near boundary conditions are well tested, in addition to average conditions or those central to the normal zone of performance.

A difference with regard to responsibility for public risk clearly exists between data processing applications of computers, with which we have more experience and familiarity, and those involving artificial intelligence and expert systems. In some instances, and according to the interpretations that prevail in various countries, this matter of responsibility may also extend to liability.

For data processing applications it is necessary, when errors occur, that there be a channel for appeal. In many instances the mere availability of a channel for appeal is sufficient. Errors usually occur infrequently, and can be accommodated, for example, if your credit balance is restored correctly next month. In applications where human lives may be involved, or when there is no opportunity to correct the error "next month", something more than "a channel for appeal" and "on-average" performance is clearly required.

#### **Regulatory Bodies**

Under somewhat analogous conditions it is not unusual for governments to establish a regulatory or administrative body to ensure that the needs of public safety, or of the individual, are met. Thus in Canada we have the Atomic Energy Control Board (AECB) for nuclear matters, the Canadian Aviation Safety Board, the Canadian Standards Association (CSA), Emergency Preparedness

Canada, the Food and Drug Act, the Canadian Human Rights Commission, and many others.

Is a similar, but special, body needed for software and engineering systems reliability in applications of a sensitive nature? This would include applications of artificial intelligence and expert systems as a special section. Given the variety of applications possible, and the difficulty of maintaining expertise in so many diverse areas, it seems reasonable to propose that this function, if needed, should be performed on a decentralized basis, that is, separately, for transportation, civil engineering structures, medical practise, and so forth. This does not eliminate, however, the possible need for a small core group to maintain specialized knowledge and to provide coordination.

Two or three decades ago the Canadian Information Processing Society (CIPS) established an ombudsman on a voluntary basis for complaints concerning data processing applications. As might be expected, it is believed that this function, while performing well, was mainly concerned with appeals concerning billing and accounting systems. In the computer applications of today, and especially of tomorrow, much more is involved.

The Canadian Standards Association has announced the availability of four voluntary standards for software quality assurance. In critical applications Q396.1.1 is for the development of software and Q396.1.2 is for "off-the-shelf" software. Correspondingly, in non-critical applications Q396.2.1 is for the development of software and Q396.2.2 is for "off-the-shelf" software. In addition, while use of the guidelines and good practise measures is voluntary, it is also hoped that organizations purchasing the standards will "register" with the CSA and agree to a voluntary audit of the degree to which the recommendations are adopted. The CSA initiative, believed to be the first software quality assurance program in North America, is also being used as a base within the International Standards Organization (ISO) (5).

#### **An Industry Association for Voluntary Accreditation of AI Products**

The extravagant claims of some commercial companies in the marketing of their AI products have created concerns amongst researchers and among the more conservative firms providing AI

products and software. In the U.K. it has been proposed that an AI industry association is needed with a code of ethics for advertising in order to protect the reputation of legitimate researchers and firms, to protect the public from claims deemed to be extravagant, and to maintain ethical standards for the AI community. If formed, the association would in effect provide a "Good Housekeeping" seal of approval for the products and work of its members.

### Codes of Ethics

Many professions, medicine, law, and engineering for example, include courses on ethics in their university curricula, and maintain a code of ethics for their members following graduation. If it has not already been done in all institutions, it would appear timely and appropriate for subject matter on ethics and professional responsibility to be added to curricula for degrees in computer science. The codes of ethics for the professions might also be examined to determine if they contain clauses that are adequate in the light of today's AI and software technology.

This alone is not enough, and frankly it can be argued that the world is not equipped for what is happening. Not only are the professions less than fully equipped for the technological change that is occurring, but a large amount of computer software, including expert systems and AI applications, will increasingly be written by persons outside the professional societies. This is only natural, and almost certain to increase. Codes of ethics developed within the professions, therefore, do not apply to these situations.

There is also an increasing amount of software in the "public domain", which raises the question, "who is responsible for the integrity of this?"

One could also raise the question as to what is meant by software in the "public domain" or what is the process for placing software in the public domain. This could be discussed at length as a separate treatise.

### Validation

Validation is an important, but particularly difficult, aspect of expert systems.

Expert systems, like all other computer programs, can contain undetected and hidden errors. These can arise as errors or "bugs" in the computer coding, or they can have their source in the background analyses and knowledge base. It is particularly difficult to fully test and debug programs that have a large number of program states. The number of branches and possible program states can in fact be extremely large. This poses a combinatorial problem so large that it is impossible to exercise and pre-test all possible conditions before use. This is one reason why some computer programs occasionally run for years before certain errors appear and are only then discovered. In addition, the nature of expert systems is such that it is frequently difficult to identify and pre-test the program over the full range of all possible input conditions.

As a further but important point, some expert systems are difficult to test definitively, because the program logic is based on inferences which have only statistical significance, and which therefore are not necessarily conclusive. Unlike engineering programs based on the laws of mechanics or physics, the output of an expert system can carry with it an element of doubt or uncertainty even when the program runs "correctly" in the computer sense. A somewhat similar condition exists with some applications of artificial intelligence other than expert systems, for example, those which use sensor-based input. There is always an element of doubt attached to the performance of the sensor, what it really measures, and how it is interfaced to the computer. This element of doubt, attached to the input, carries through to the program output.

There is an additional doubt with expert systems that the knowledge base may not contain all the necessary rules at any particular time or stage of its development, or that the computerized search through the knowledge base may not have considered all possible paths.

These are some of the reasons why expert systems, unlike other computer applications, are required to serve a period of apprenticeship during which their behaviour is studied in the presence of experts. During this time period confidence in the systems capabilities can gradually rise as positive experience is gained and as some problem areas are corrected. At the same time knowledge

is obtained of weaknesses and deficiencies in areas to be avoided until improvements are made.

### Traceability

One of the attributes of expert systems is that they should be able to explain to their user how a particular result was obtained. That is, to explain or reveal the logic or reasoning path that was used. Similarly if additional input is requested from the user, the program should be able to explain, on interrogation, why it is needed or how it will be used. This attribute of traceability, or transparency as it is sometimes called, is important to the user in understanding how the program is using the input provided and accordingly how it is reaching its conclusions. This is a particularly important attribute in instances where the user must take professional responsibility for the advice given by the expert system.

It should be noted, however, that the presence of this attribute alone, or by itself, does not in all cases completely shift the burden of responsibility for correctness from the program developer to the user. It is of little value, for example, in a realtime expert system sorting out hundreds of alarm signals in seconds during an upset in a large chemical plant to reveal days or weeks after the wrong action was taken that this would have been apparent to the operator at the time if the traceability or transparency capability had been used.

### Documentation

Somewhat similar remarks can be made concerning program documentation. If a user is to take full responsibility for an expert system, he must have system documentation available to him that permits him to achieve the level of understanding necessary for this. This poses a special requirement on the user level documentation for expert systems. In a world where even the documentation for many conventional computer programs and software packages does not achieve this level it is a cause of concern.

Not only must the expert system user achieve a very high level of understanding, but also it is necessary to ask if he or she will be able to do so with a reasonable level of effort expended over a reasonable period of time. It is to be expected that the larger and more complex systems will

only be adequately understood at special centres of excellence or locations close to their origin and sometimes only by their originators. This is where questions related to knowledge sovereignty and a possible "software colonialism" arise, as mentioned elsewhere in this report.

If a deep and substantial knowledge is required by the capable user of a large system, the knowledge and education requirements for high level expert system developers is even greater. It is conceivable that a lifetime of hard work and institutional support may be necessary in the future in many instances.

The statement that good program documentation is necessary seems to imply that good documentation alone will solve the problem. This is not the case. It can take years of experience and close association with a computer program to fully understand the implications of its output. Expert systems are an example of this, and as expert systems increase in their depth of reasoning, this learning time for the user will increase. Good documentation is necessary, but it is only a starting point.

### Summary and Conclusions

While codes of ethics, professional responsibility, and regulatory bodies are useful mechanisms, they are by no means likely to be sufficient. Computer software, including expert systems, will soon be almost as common as people. It will originate from a wide variety of sources, and from people in many countries. Only a small percentage of these sources will fall within those covered by professional codes of ethics, in terms of either software generation or use. Thus, it is fair to say that, if some form of regulation on supply or use is needed, it will be extremely difficult to establish and achieve.

Caveat emptor, let the buyer beware, might appear as another alternative, but is equally non-viable. Software is already being given away as an inducement to the purchase of other goods and services, and this practise will increase as software becomes more and more a low cost commodity item. In many instances the buyer (or acceptor) of software will lack the skills and knowledge for any in-depth evaluation or validation of software he acquires.

Should some form of control become necessary, there is virtually no workable method by which it could be implemented that would not permit circumvention by those who might wish to do so. It appears that in most instances this could be quite readily accomplished. In the meantime, the current system rests on:

- the integrity of software developer(s), wherever they may be;
- the codes of ethics of professional societies and other bodies, in instances where they apply;
- regulatory codes for applications already recognized to be sensitive to public safety, e.g., nuclear power plant control;
- caveat emptor;
- education and awareness of the general public.

---

### *Section III*

#### *A Protocol for the Examination of AI Applications from the Social Point of View*

### ***Section III — A Protocol for the Examination of AI Applications from the Social Point of View***

The purpose of this protocol on the bringing into use of a system employing artificial intelligence is to alert to the sociotechnical dimensions of doing so the persons who:

- define the ground for the intended uses of the system,
- create the system,
- bring it into use,
- operate it,
- have their tasks influenced by its use,
- evaluate its worth.

The importance of reflecting on these sociotechnical dimensions lies in the reality that systems of artificial intelligence have the potential to carry out significant tasks of perception, reasoning, decision making, and taking action that have characteristically been performed by persons. This potential raises new and profound aspects of questions about the relation of technological change to work, organization, and culture. Here the word culture is taken to mean the whole evolving social ecology which forms the basis on which the members of a society interpret their experiences and shape their relations with one another and their environment into a way of life. This ecology is multi-dimensional and encompasses the social character of science, engineering, technology, economics, law, politics, the arts, morality, and religion.

#### **Contexts for a Case Study**

In the uses of artificial intelligence, as indeed of all technology, a fundamental cultural question can be put in two extreme ways.

- First, are persons to be viewed as being required (by whom under what authority) to adapt to technological artifacts?
- Second, are artifacts to be viewed as being brought into use to complement, preserve, and assist the expression of (whose) human talents?

Since technological change can and must be considered from the differing perspectives of individuals, groups, organizations, and indeed society there is clearly a spectrum of answers lying between the limiting questions as stated. For this

reason it is imperative in case studies that ecological questions of intent, expectations, and consequence be considered at a number of different contextual levels which are selected by consciously shifting the focus of attention by a process of what may be called "cognitive zooming".

#### **The Focal Organizational Context**

The organizational setting within which an application of artificial intelligence is embedded constitutes a focal context from which attention can be directed inward to operational structure and can be directed outward to various social contexts. The embedding of a use of artificial intelligence is an innovation in the organization. This innovation represents:

- the set of all considerations and actions associated with the bringing into practical operational use the system embedded in the organizational setting,
- the consequences to the organization and its people.

Just as the innovation of an industrial product such as a new material, device, or system does not typically follow a linear sequential process from basic science, to applied science, to engineering design, to production, and to market distribution, the innovation of an application of artificial intelligence in an organization will not proceed simply from prior specification of intended use, to system design, to implementation and evaluation. The interactions between those who require, specify, design, construct, install, document, operate, manage, and evaluate will typically be complex. These interactions will reflect the evolving interplay of social and economic purposes expressed through operational practices within and outside the focal organization setting.

#### **\*IDENTIFY**

- with respect to the selection of an application for embedding, identify at both the corporate and relevant social levels distinctive contextual characteristics.

For example, a private-sector industry has the underlying corporate purpose of making profit to provide a reasonable return on investment to the owners. This objective is carried out within an environment of corporate law and regulation, which together with socioeconomic circumstances,

local, regional, and global, establish a context of risk and competitive opportunity for the responsible achievement of profits.

**\*IDENTIFY**

- the distinctive socioeconomic circumstances of the organizational setting (such features as the form of the corporate entity, its relation to organized labour, to the local and wider environment of competition and regulation, to local social, political, and ecological conditions).
- within the organizational setting identify the motivation for bringing into use the application. Is there a perceived problem to be addressed? Is the instance part of a corporate strategy for technological change? Is there an internal champion as a driving force? Is there persuasion from outside suppliers? Is there peer pressure from the industrial sector?

**\* IDENTIFY**

- within the concrete situation of application the proposed grounds of justification for the necessary expenditure and stress of organizational change? In an industrial setting these elements will characteristically have to do with technical feasibility, profitability, and quality of human relations. Factors that may be pertinent include potential for increased productivity, improved product quality, improved working conditions and job satisfaction, reduced downtime, enhanced flexibility of production and organizational responsiveness, reduced costs or energy, materials, training, labour, and management.

To achieve reasonable clarity of response to the foregoing question of justification it is necessary to examine how the application of artificial intelligence will be embedded into and change the existing organizational context.

**\*IDENTIFY**

- the organizational changes necessary to accommodate the system. These changes will include redefinition and reassignment of operational tasks and the associated communications network and structure of accountability. Such redefinitions will account for job creation, job displacement, machine and process replacement, and identify as enhancements or reductions significant shifts in the scope of both operational and managerial tasks. These

definitions of organizational change will also identify anticipated changes in the degree of centralization/decentralization of managerial control and of worker participation.

Since current systems of artificial intelligence, for example in the form of expert systems, have sharply limited and specialized domains in which they can perceive, question, reason, make decisions, and take actions in the face of uncertainty, it is essential that the relation of its operation to human oversight be understood by all persons responsible for the operations and consequences of the employment of the system. This requirement raises novel issues of task definition, corporate accountability, and legal liability.

**\*IDENTIFY**

- the organizational and operational contexts and characteristics of the system in terms of the functions of perception, questioning, being questioned, reasoning and explaining reasoning, decision making, and taking action through effectors.
- the boundaries or limits of the domain of expertise represented in the system and the manner in which human oversight and accountability for the use of the system is to take place in the organizational context. The designers, suppliers, and the domain experts employed have a distinctive responsibility to document and explain to the using organization "the micro-world" within which the system is able to function with the intended reliability.

These identifications will serve to define the extent to which the system operates to fulfil complete tasks or functions in an advisory or enabling manner with respect to the human operation and management of the organization.

There is substantial evidence to show that technological and therefore organizational change is most effectively carried out and evolved when those persons participating in and affected by the change are kept informed of the contextual elements which their experience, skills, and talents enable them to grasp and potentially contribute to.

**\*IDENTIFY**

- the practices to be followed in the organizational context:

- for keeping informed from the earliest stage of system choice, and "employing" the capabilities of those who will participate in and become part of the innovation,
- to inform and assist in a timely manner persons whose customary jobs will cease to exist,
- to document evolutionary changes in design and operation,
- to provide training in the use of the system.

Whether or not systems of artificial intelligence will realize the expectations held out for them in enhancing the performance of the organizations in which they become embedded as innovations will become gradually known only through a holistic process of evaluation against as many contextual elements of the embedding process as can reasonably be examined.

Systems of artificial intelligence represent a unique phase of technological change in which the consequences of use will gradually reverberate more deeply into the form and ethos of the organizational setting than has any previous stage of technology. Sensitivity to questions of change of organizational ethos is an essential dimension

in the evaluation of consequences of use. As the characteristics of those organizational settings undergo change, the consequences will reverberate into the whole social ecology of our culture. The background documents for this protocol consider key dimensions of such reverberation.

**\*IDENTIFY**

- the practices intended to be followed to assess the contribution (or otherwise) of the system to the performance and overall welfare of the organization using technical, organizational, financial, and socially germane norms. These practices will include taking note of changes in ethos reflected, for example, in internal social networks among participants (workers and managers), and seek to understand changing attitudes to the place and significance of work in an organizational context wherein the exercise of perception, reasoning, and decision making is distributed among persons and machines.

The latter assessment will require accumulation of understanding from the larger context of sets of similar applications characteristic of the development of the field of artificial intelligence and a sensitivity to the evolving social ecology of technological change.

*Section IV*

---

*Summary and Recommendations*

### **Section IV — Summary and Recommendations**

Applications of artificial intelligence should not be pre-judged, out of context. This means that they can only be examined within their own organizational and unique setting. Accordingly, this report, while recommending that planned or existing applications can usefully be examined from the social as well as from the business and technical point of view, does not purport to pre-judge applications, but does present a protocol by which such examination can be made. Central to the protocol is the belief that technological systems can have influence or impact at the direct level on users and workers and former workers; at the intermediate level on organizations and; at the macro level on nations and human society at large. While artificial intelligence may or may not represent a significant departure from past forms of technological development (and information systems in particular), many of the criteria for evaluation derived by experience from these earlier forms, apply also to artificial intelligence. It appears likely in the minds of many that artificial intelligence carries with it some concerns that are new in kind, or that are at least of heightened magnitude from those of the past. For example, a fundamental and singular thing about AI is that it is changing the way in which the process of reasoning and decision making is proportioned between people and machines. Questions related to the professional responsibility for expert systems, and perhaps also for responsibility and liability in the legal sense particularly when mistakes occur, arise from these sources.

As with any technology, there are both pros and cons associated with it, and particularly with regard to the way in which it is used. Among those experienced with artificial intelligence, the following have been noted by some organizations to date:

#### **Pros**

1. increased job satisfaction
2. quality of work improved
3. VDUs (video display units) are less noisy and more flexible than the mechanical devices they replace
4. greater response and ability to be on top of affairs
5. new systems are user friendly, simpler to learn

6. computer solution is often cheaper than mechanical means

#### **Cons**

1. job security may be threatened
2. physical strain, mental strain, and possible health hazards associated with the use of VDUs
3. fear of innovations, loss of value of "experience"
4. fear of change, inability to cope with non-permanence
5. technology introduction requires planning, education, and attention to details
6. need for more flexible, more educated mind-set
7. loss of need for "specialist knowledge"
8. shortage of specialists
9. utter simplicity — job de-skilling and lack of challenge.

Fundamental to almost all the considerations posed is a changing balance in how decision making is conducted and performed. This is a change in balance between the portion delegated to machines, and the portion reserved for people as human beings. This changing human/machine symbiosis is viewed by many as a fundamental point of departure in the development of technology. Since the beginning of the industrial revolution, large changes have been witnessed in the division of labour between people and machines. We now, potentially, have the beginning of a new era with a significant change in the division of reasoning and decision making between people and the computer. The outcomes of this are likely to be profound.

#### **Summary of the Protocol**

1. Are persons required to adapt to the artifacts of the system, and if so, under what authority?
2. Will the system, as brought into use, preserve and assist the expression of human talents? If so, whose talents will be so preserved and assisted?
3. Examination of the system should identify:
  - The characteristic purposes of the organization at both the corporate and relevant social level
  - The distinctive socio-economic circumstances of the organizational setting (i.e., form of the

corporate entity, relationship to organized labour, competitive environment, regulatory environment, etc.)

- The motivation(s) for introducing the application. (e.g., is it to solve a perceived problem, is there an internal champion or external advocate such as a supplier?)
- The justification(s) for the application, the expenditures necessary, and the organizational changes necessary.
- How the system will be embedded into the existing organizational context, and how it will change that context.
- The organizational changes necessary to accommodate the system. This will include job creation, job displacement, machine and process replacement, the identification of enhancements, redefinition and reassignment of operational tasks, changes in the communications network, changes in accountability, changes in the degree of centralization of managerial control and of worker participation.
- How the system will function in the face of uncertainty and the degree of human oversight required.
- The organizational operational contexts and characteristics of the system in terms of the functions of perception, questioning, being questioned, reasoning, explaining reasoning, decision making, and taking action through physical effectors.
- The boundaries or limits of the domain of expertise represented in the system.
- The manner in which human oversight and accountability for use of the system is to take place in the organizational context.
- "The microworld" within which the system is able to function with the anticipated reliability. This will serve to identify the extent to which the system operates to fulfil complete tasks or functions in an advisory or enabling manner with respect to the human operation and management of the organization.
- The practices to be followed in the organizational context:

- for "employing" the capabilities of those who will participate in and become part of the innovation and for keeping them informed from the earliest stage of system choice
- to inform and assist in a timely manner persons whose customary jobs will cease to exist
- to document evolutionary changes in design and operation
- to provide training in the use of the system
- To the extent possible, how the consequences of use of the system will reverberate into the form and ethos of the organizational setting and eventually into the culture of society itself.
- The practices intended to be followed to assess the contribution (or otherwise) of the system to the performance and overall welfare of the organization using technical, organizational, financial, and socially germane norms. These practices will include taking note of changes in ethos reflected, for example, in internal, social networks among participants (workers and managers), and seek to understand changing attitudes to the place and significance of work in an organizational context wherein the exercise of perception, reasoning, and decision making is distributed among persons and machines.
- The latter assessment will require accumulation of understanding from the larger context of sets of similar applications characteristic of the development of the field of artificial intelligence and a sensitivity to the evolving social ecology of technological change.

#### Additional Issues, Concerns, and Questions to Ask

A list of issues and concerns to be considered when applying an AI system and/or when reviewing a case study should also include:

#### *How and why has the application been chosen?*

- Different reasons for implementing an AI system will result in very different social consequences. For example, was the application chosen to reduce manpower, to increase

productivity, to free up manpower for other activities, or for training purposes?

- Was there user involvement in the choice and design of the application? Did the AI system benefit from the user's understanding, advice and creative abilities?

#### *How does the AI System fit into operations?*

- Does it address a problem that is recognized as a problem by the users?
- Are there organizational changes necessary to accommodate the system and if so how will they be addressed?
- Will implementation of the system introduce or cause changes in the communication patterns within the organization?
- Will there be job displacement? Will there be job creation?
- Will there be productivity gains? How, and to whom, are the benefits of these to be distributed?
- When and how is the technology introduced to staff (both to the domain experts and users)?
- What are the initial expectations about the technology prior to implementation? How is the technology to be evaluated after implementation?
- What level of the organization is most directly effected by the technology? management? workers?
- What is the effect of the system on the user's job? Is the user's job enhanced or degraded? Do the users have more or less time for other aspects of their jobs? Do the users make better or worse use of their time? Is their performance improved or degraded? Is their job made more of less tedious? Is their job made more or less difficult?

- What are the costs and benefits as seen by the domain expert involved in building the system?
- What are the costs and benefits as seen by the user of the system?
- What will the consequences be in terms of the centralization or decentralization of the

organization or in terms of changes in management; increase/decrease in flexibility/rigidity of responses?

- Is the user interface well designed?

*Recognizing that the system, when dealing with uncertainty, may give erroneous or incorrect output in the form of advice, conclusions, or direct action, has the responsibility for this been clearly defined?*

- Is the responsibility chain back to the originators of the system components clearly defined? For example, are the domain experts, knowledge engineers, system engineers, and computer scientists involved identified or identifiable by name?
- Does use of the system involve a risk to the safety and health of the public at large, or to individual members of the public? Could use of the system result in personal damage to a group or to an individual?
- How is responsibility for use of the system allocated and shared between persons or bodies such as the originators of the system components, the practitioner who is operating the system, the end user on whose behalf it is being utilized? Where necessary, is this allocation of responsibility clear in the legal sense?
- If all, or part of, the system input is sensor based, how is the system protected against false, erroneous, or incomplete input or errors in the meaning and interpretation of the input? How are such conditions alerted to the user? Is the system "fail safe" from the user point of view if the system is suddenly interrupted and made unavailable?
- Does the domain of the system fall within the domain of one of the professions, such as engineering, medicine, or law? Does its use comply with the regulations and the codes of ethics of the professions concerned?
- Does the domain of the system lie within the bounds of other regulatory bodies such as the transportation act, food and drug act, environmental regulations, health and safety, the handling of hazardous wastes, etc.? If so, has compliance with those regulations been considered and assured?

- Recognizing that use of the system may improve performance "on-average", but not necessarily in all circumstances, what provision exists to screen out or minimize those instances in which degraded results will occur?
- In instances where human intervention should, or must, be applied to the system output, will the operator or user of the system have reasonable time in which to apply this intervention? This is particularly important in the case of real-time systems.
- In instances where false or undesirable information or actions may result from use of the system, does a reasonable channel of appeal exist for those who may be affected?

#### Documentation

- Is documentation complete and available to the user, permitting him to fully understand and comprehend the "micro-world" within which the program operates?
- Does this documentation clearly define the limits and limitations of the program?
- Can the user reach the level of understanding necessary to use the program safely and properly with a reasonable amount of study and effort over a reasonable period of time? What level of prior education and experience does this assume?
- Can the system be accessed and used by persons without this level of capability?

#### Validation

- Is the system validation considered complete, or is the system for approved experimental use only? What special safeguards does this require?
- Has the system validation considered all necessary possibilities?
- What is the source of domain knowledge or analysis as assumed and used in the system? Is this domain knowledge or analysis description included in the documentation? Is this in terms the user can understand? Are the assumptions and limitations clearly defined?

#### Traceability

- Can the system readily and adequately explain how any given result was obtained, and identify the information which was used or not used?
- Can the system readily and adequately explain how any item of input normally required, or requested, will be used?

#### Updating

- What provision has been made for reviewing the system, keeping it current and up to date? How are users informed of these changes?

#### Privacy

- Does normal use of the system affect the privacy of any individual? Could this result through abnormal use, or in combination with some other system?

#### National Control

- Does use of the system remain within the national boundaries of the country of origin? In cases where usage is considered outside such boundaries, does this result in a loss of national control in the social or economic sense?

#### Ethical

- Are there concerns or issues of an ethical nature associated with the system or its use?

The above are issues and concerns, largely expressed in the form of questions that should be explored when considering new AI technology or when reviewing an existing application of the technology.

The concerns and the ways in which they are expressed may be different in subtle but important ways from the larger population of interest, i.e., those groups implementing AI technology into their organizations. It may be noted that procedures and methods for generating questionnaires designed to elicit the concerns and issues from the population of interest are available that avoid the pitfall of assuming that the evaluators are representative of the population of interest.

#### Summary of Issues and Concerns

1. A realistic examination should be attempted of current and future prospects for AI applications at home, in the workplace, and in the government. The discussion and protocol for

such examination contained in this report is intended to encourage and facilitate this process.

2. Such examination would include the impact of computer-related technology in the workplace, balancing benefits against perceived problems, including deskilling, monitoring, job loss, restricted promotion paths, breakdown of traditional social organizations in the office, limited entry level opportunities, and health-related concerns.
3. The implications of partially realized intelligent systems in terms of the requirements placed on humans to accommodate to system inadequacies must be considered. In the haste to introduce AI into the workplace, pressures may be placed on people to work with systems, which, while advertised as intelligent, are seriously deficient in many areas.
4. Of particular interest is the role of AI in decision making, whether in financial institutions, in the executive suite, or in diverse military situations.
5. Intelligent systems may find ready application in intelligence activities such as the automatic interpretation of tape recordings and the cross-correlation of electronic files. Added to current threats to privacy, the availability of such powerful mechanisms could increase real and anticipated assaults on individual privacy.
6. Futuristic projections of a society without poverty brought about by the extensive diffusion of AI applications have been considered by science writers and more recently by AI researchers themselves. Speculation is interesting but the assumptions underlying the forecasts must be carefully analyzed. Questions to be considered include the following:
  - What replaces regular work as a necessary part of life?
  - How is wealth to be distributed if a wage system is no longer operative?
  - How will the political structure respond?
  - How will human dignity and self worth be affected if we are no longer defined in part by what we do?

It should be kept in mind that these questions are obviously so difficult to answer, or even to

characterize, that only a beginning is made here. However, it is important that they be raised and that a serious discussion be initiated.

#### Conclusions

- Artificial intelligence has a great potential to contribute to the well-being of humanity in both the economic and social sense.
- Research in artificial intelligence may contribute, in consort with parallel research by other disciplines, to the understanding of the human mind and even to the understanding of human behaviour.
- Applications of artificial intelligence have the potential to greatly extend the boundaries for current computer applications into regions involving recognition, reasoning, problem solving, and learning. Applications include natural language processing, machine translation, diagnostic systems, and autonomous systems, including robots, with sensing and logic capabilities permitting them to respond to changes in their external environment.
- The range of application of expert systems, which is one branch of artificial intelligence, is particularly broad, reaching into almost every field of human endeavour in business, industry, education, health care, recreation, the arts, defence, and war.
- Aside from the direct applications of artificial intelligence, research in artificial intelligence may contribute to the understanding of how people think, which includes logically, creatively, and irrationally, and to how people learn. The understanding gained may lead to improvements or enhancements of these processes.
- At the same time, artificial intelligence may alter the way in which man thinks of himself. Some see man's exposure to his weaknesses as a beneficial process. With others there is a concern that by apparently reducing thinking to a mechanistic process, the public perception may be that people are regarded as "nothing but clockwork". This may have "indirect effects on self esteem and social relations [that] could be destructive to many of our most deeply held values" (6).
- The most distinctive and singular thing about AI is that the process of reasoning used to

make decisions becomes fundamentally altered in the way it is distributed between people and machines.

- Because people may be directly affected by artificial intelligence applications, and by expert systems in particular, either individually or in some instances in large numbers, there is a special need to assure the quality, verification, safety, and reliability in a wide variety of applications.
- At the present time there does not appear to be any specific mechanism for this in Canada, or elsewhere, other than the professional responsibility of individuals and the codes of ethics that the professions have developed for their members.
- The diversity of applications for artificial intelligence suggests that this decentralized approach to responsibility may be the best to follow, but with special attention drawn to the responsibilities involved and an examination of the codes of ethics, where available, to ensure that these new conditions are adequately represented.
- To assist in the examination of artificial intelligence applications in the social context, a protocol for the examination of case histories or proposed applications has been developed and is presented herein.
- Many other new technologies are currently being introduced besides artificial intelligence. Some of these, such as advanced manufacturing systems are likely to precede artificial intelligence in terms of a very broad scale implementation, and with very widespread social and economic effects. The final impacts of artificial intelligence, whatever they may be, will occur in the presence of these parallel activities, some of which are known, and also in the light of other events which are not known or predictable.

### The Industrial Revolution, Revisited

Relax. Don't Worry. ~~Machines will do the~~ ~~Computers~~ ~~Thinking~~ ~~Work.~~

## Section V

### References

**Section V — References**

1. Scrimgeour, J., Expert Systems — A Review of their Potential and Application in Canada, *Canadian Conference on Industrial Computer Systems*, May 1986 (principal source).\*
2. Peacocke, R., Zlatin, D., Artificial Intelligence at Bell-Northern, *Artificial Intelligence (Expert Systems) Symposium*, NRC, January 27, 1987 (with adaptation).
3. Rosenberg, R., The Impact of Artificial Intelligence on Society, Department of Computer Science, University of British Columbia (manuscript submitted for publication, August 1988).\*
4. *IEEE Spectrum*, September 1988 p. 3.
5. CSA Guidelines Address Quality Assurance for Software Developers, *Direct Access*, January 27, 1989.
6. Boden, M.A., Artificial Intelligence. *The Oxford Companion to the Mind*, 1987.

---

\*Reviewed in the Biblioabstracts Section, Appendix 1.

*Appendix 1*

---

*Terms of Reference: NRC Associate Committee  
on Artificial Intelligence*

**Appendix 1 — Terms of Reference:  
NRC Associate Committee on  
Artificial Intelligence**

**Subcommittee on the Social Context of  
Artificial Intelligence**

**Objectives and Activities**

1. To conduct a brief literature search and prepare a bibliography of references pertinent to the scope of the sub-committee.
2. To develop a brief report or working paper to identify typical issues and areas of concern.
3. To develop a short list of sub-committee members and a larger network of interested individuals with knowledge and expertise in the subject area.
4. To hold at least one meeting or consultative workshop to review material available, areas of concern, available expertise, and develop guidelines for the avoidance of problem areas.

Adopted: February 16, 1988

**Members**

- Dr. R. De Mori, CRIM & McGill University, Montreal, P.Q.\*
- Professor C. Gotlieb, University of Toronto, Toronto, Ont.†
- Dr. J. Ham, University of Toronto, Toronto, Ont.‡
- K. Hayes, Canadian Labour Congress, Ottawa, Ont.†
- Dr. C. Lajeunesse, Centre de recherche informatique de Montréal (CRIM), Montreal, P.Q.†
- Dr. A. Mackworth, University of British Columbia, Vancouver, B.C.†
- A. Mayman, National Research Council of Canada, Ottawa, Ont.†
- Professor R. Rosenberg, University of British Columbia, Vancouver, B.C.†
- J. Scrimgeour, National Research Council of Canada, Ottawa, Ont.†
- G.F. Sekely, Canadian Pacific, Toronto, Ont.†
- Dr. B. Smith, Acquired Intelligence, Victoria, B.C.†
- G. Thomas, Price Waterhouse, Ottawa, Ont.\*
- Professor W. Vanderburg, University of Toronto, Toronto, Ont.†

\* As of October 1988.

† As of May 1988.

‡ Chairman.