THE RELIABILITY OF EXPERT SYSTEMS

ELLIS HORWOOD BOOKS IN INFORMATION TECHNOLOGY

General Editor: Dr JOHN M. M. PINKERTON, Principal, J & H Pinkerton Associates, Surrey (Consultants in Information Technology), and formerly Manager of Strategic Requirements, ICL

M. Barrett & A. C. Beerel — EXPERT SYSTEMS IN BUSINESS: A Practical Approach

M. Becker, R. Haberfellner and G. Liebetrau — ELECTRONIC DATA PROCESSING IN PRACTICE: A Handbook for Users

A. C. Beerel — EXPERT SYSTEMS: Strategic Implications and Applications

K. Bennett — SOFTWARE ENGINEERING ENVIRONMENTS: Vol. 2

A. Bradley — OPTICAL STORAGE FOR COMPUTERS: Technology and Applications

P. Brereton — SOFTWARE ENGINEERING ENVIRONMENTS

R. Bright — SMART CARDS: Principles, Practice and Applications

D. Clarke and U. Magnusson-Murray — PRACTICAL MACHINE TRANSLATION

V. Claus and A. Schwill — ENCYCLOPAEDIA OF INFORMATION TECHNOLOGY

D. Cleal — KNOWLEDGE-BASED SYSTEMS: Implications for Human–Computer Interfaces

I. Craig — THE CASSANDRA ARCHITECTURE: Distributed Control in a Blackboard System

T. Dailer, *et al.* — SECURITY OF INFORMATION AND DATA

D. Diaper — KNOWLEDGE ELICITATION: Principles, Techniques and Applications

G. I. Doukidis, F. Land and G. Miller — KNOWLEDGE-BASED MANAGEMENT SUPPORT SYSTEMS

P. Duffin — KNOWLEDGE-BASED SYSTEMS: Applications in Administrative Government

C. Ellis — EXPERT KNOWLEDGE AND EXPLANATION: The Knowledge–Language Interface

A. Fourcin, G. Harland, B. Barry and V. Hazan — SPEECH INPUT AND OUTPUT ASSESSMENT: Multilingual Methods and Standards

M. Greenwell — KNOWLEDGE ENGINEERING FOR EXPERT SYSTEMS

F. R. Hickman — ANALYSIS FOR KNOWLEDGE-BASED SYSTEMS: A Practical Guide to the KADS Methodology

P. Hills & J. Beaumont — INFORMATION SYSTEMS: Management Implications for the Human Computer Interface

E. Hollnagel — THE RELIABILITY OF EXPERT SYSTEMS

J. Kriz — KNOWLEDGE-BASED EXPERT SYSTEMS IN INDUSTRY

M. McTear & T. Anderson — UNDERSTANDING KNOWLEDGE ENGINEERING

W. Meyer & H. Walters — EXPERT SYSTEMS IN FACTORY MANAGEMENT: Knowledge Based CIM

U. Pankoke-Babatz — COMPUTER BASED GROUP COMMUNICATION: The AMIGO Activity Model

J. M. M. Pinkerton — ADVANCED INFORMATION TECHNOLOGY

S. Pollitt — INFORMATION STORAGE AND RETRIEVAL SYSTEMS: Origin, Development and Applications

S. Ravden and G. Johnson — EVALUATING USABILITY OF HUMAN–COMPUTER INTERFACES: A Practical Method

P. E. Slatter — BUILDING EXPERT SYSTEMS: Cognitive Emulation

H. M. Sneed — SOFTWARE ENGINEERING MANAGEMENT

M. Stein — BUILDING EXPERT SYSTEMS MODELS FOR DATA TRANSMISSION

R. Stutely — ADVANCED DESKTOP PUBLISHING: A Practical Guide to Ventura version 2 and the Professional Extension

J. A. Waterworth and M. Talbot — SPEECH AND LANGUAGE BASED COMMUNICATION WITH MACHINES Towards the Conversational Computer

R. J. Whiddett — THE IMPLEMENTATION OF COMPUTER SYSTEMS

THE RELIABILITY OF EXPERT SYSTEMS

Editor
ERIK HOLLNAGEL
COMPUTER RESOURCES INTERNATIONAL A/S
Advanced Information Processing Division

ELLIS HORWOOD LIMITED
Publishers · Chichester

Halsted Press: a division of
JOHN WILEY & SONS
New York · Chichester · Brisbane · Toronto

First published in 1989 by
ELLIS HORWOOD LIMITED
Market Cross House, Cooper Street,
Chichester, West Sussex, PO19 1EB, England
The publisher's colophon is reproduced from James Gillison's drawing of the ancient Market Cross, Chichester.

Distributors:
Australia and New Zealand:
JACARANDA WILEY LIMITED
GPO Box 859, Brisbane, Queensland 4001, Australia
Canada:
JOHN WILEY & SONS CANADA LIMITED
22 Worcester Road, Rexdale, Ontario, Canada
Europe and Africa:
JOHN WILEY & SONS LIMITED
Baffins Lane, Chichester, West Sussex, England
North and South America and the rest of the world:
Halsted Press: a division of
JOHN WILEY & SONS
605 Third Avenue, New York, NY 10158, USA
South-East Asia
JOHN WILEY & SONS (SEA) PTE LIMITED
37 Jalan Pemimpin # 05–04
Block B, Union Industrial Building, Singapore 2057
Indian Subcontinent
WILEY EASTERN LIMITED
4835/24 Ansari Road
Daryaganj, New Delhi 110002, India

British Library Cataloguing in Publication Data
Hollnagel, Erik
The reliability of expert systems. —
(Ellis Horwood books in information technology)
1. Expert systems. Reliability
I. Title
006.3'3

Library of Congress data available

ISBN 0–7458–0541–8 (Ellis Horwood Limited)
ISBN 0–470–21502–X (Halsted Press)

Printed in Great Britain by The Camelot Press, Southampton

Table Of Contents

VII ISSUES IN THE RELIABILITY OF EXPERT SYSTEMS

PREFACE

This book is the result of a one-day seminar on 'Safety and Risks in the Use of Expert Systems' which was held in Copenhagen on May 19th, 1988. The idea to have a seminar about this topic evolved from the discussions in the 'Sub-committee for Computer Systems' working under the auspices of the 'Committee on Risk Assessment'. The 'Committee on Risk Assessment' was appointed by a resolution in the Council of the Danish Academy of Technical Sciences in February 1985 and started its work in June the same year - so far for a period of three plus one years - with the following terms of reference:

- The object of the Committee is to promote risk assessment research and the application of its results in Denmark.
- During fulfilment of this object, the Committee must initiate and coordinate research projects and coordinate the exchange of information between creators and users of the results of risk assessment research within all professional disciplines, including ensuring improved information about international risk assessment research in Denmark.
- The Committee is primarily meant to act as promoter and coordinator of activities carried out under other auspices.

During the discussions in the sub-committee about the risk involved in the use of computer systems, it became clear that expert systems as a new type of computer systems might also represent a new, and potentially very large, source of risks. The background for the seminar was a discussion about how expert systems have changed from being a subject for research to the position where they are in practical use as advice and support systems in industry and commerce. The essence of an expert system is the ability to give advice and support in cases where the problems are incompletely understood and very unstructured. This naturally raises some concerns about the safety and reliability of their functioning. At the time there seemed to be little in the way of an established position on how to answer these questions, and the seminar was therefore organised to provide an overview of the main issues and current trends.

The seminar was organised into two sessions which focussed on the theoretical and practical aspects of safety and risk in the use of expert systems. Each session had two invited speakers, and the book presents the extended version of these four prepared

contributions, as well as an introduction to the subject and a summary based on the discussions that took place during the seminar.

Acknowledgement is due to a number of people. First of all, the four invited authors who accepted the challenge of presenting their view on Safety and Risk in the Use of Expert Systems and who afterwards took the trouble together with their co-authors to revise, and considerably extend, their presentations. One of them, Robert J. Taylor, also served as a member of the organising committee - in addition to being the driving force in the sub-committee on computer systems. The other members of the organising committee were Professor Palle Thoft-Christensen, Aalborg University Center, and the scientific secretary of the 'Committee on Risk Assessment', Ms. Birthe Schouby, who also took care of the practical arrangements of the seminar. Thanks are due to Dr. Jens Langeland-Knudsen who together with Professor Thoft-Christensen chaired the sessions, and to Mr. Pierre Laraignou, who made sure that the presentations and question & answer sessions were properly tape recorded, and who later did a good job in transcribing parts of the recordings. And I would finally like to thank the 'Committee on Risk Assessment' with its chairman, Mr. Niels Hjort, who from the beginning fully supported the idea of organising a seminar on this topic.

Erik Hollnagel

Copenhagen, December 18th, 1988

I: INTRODUCTION

Jens Langeland-Knudsen
Computer Resources International
Vesterbrogade 1A, DK-1620 Copenhagen V, Denmark

I am very pleased to get this opportunity of saying a few words to open this seminar. In looking at the topics on the agenda I remembered a short communication, by E. Dijkstra, a well known computer scientist, which I read several years ago. In this note he provided a characterisation of a number of different computer languages, which I would like to quote from memory. Generally, he was quite displeased with most of the existing languages. About BASIC it was said that teaching it to anyone should be considered a criminal offence. FORTRAN, in its turn, was compared to an infantile disorder. And, of particular interest for today's topics, was the characterisation of PL-1 which was said to belong more to the problem set than to the solution set.

Turning towards the topic of today one may ask the same question about expert systems, i.e. whether they belong more to the problem set than to the solution set. Or to put it in a different, and more positive, way, one may ask whether the benefits of using expert systems will justify the added complexity that they introduce. There should be no doubt that I believe expert system are a great benefit in many ways. One must, however, also accept that expert system are complex beasts, at least when you are looking at the life size expert system that are applied in real tasks rather than the 'toy' expert system that academics are so fond of. Another question one may ask is whether expert system are mechanical devises like other human tools, or whether there is more in expert systems than meet the eye. This is essentially the fundamental question that has always been intriguing people, scientists or lay people alike, looking at expert system. In other words, do these systems contain intelligence, and what may this intelligence be?

Taking a look at the invitation to the seminar I noticed that at least some of the organisers appear to believe that expert systems might be something more than mechanical tools. This is clear from some of the questions that are to be discussed today, i.e. whether the advice of an expert system possibly can be dangerous, whether the advice is good, whether it takes account of side-effects, traps and false assumptions, etc. In my view these would be very strange requirements to put to a normal 'mechanical' type of system. (By mechanical I do not mean mechanical in the physical

sense, but rather in the sense that normal software is mechanical.) So to me, at least, this indicates that expert systems might somehow be different from other kinds of systems.

I believe that this makes evaluating the risks and potential benefits of expert system more difficult than, say, for the kind of (mechanical) systems that we otherwise deal with. But unfortunately, or may be fortunately, I am not an expert in expert system. I would therefore like to leave the word to the invited speakers, who will tell us what problems of safety and risks there really are in the use of expert system.

II: THE RELIABILITY OF EXPERT SYSTEMS

AN INQUIRY OF THE BACKGROUND

Erik Hollnagel
Computer Resources International A/S
Bregnerødvej 144, DK-3460 Birkerød, Denmark

1. INTRODUCTION

The title of this book is 'The Reliability Of Expert Systems'. This title can be interpreted in several ways. Considering the technical use of the term reliability in e.g. reliability analysis or statistics the title can refer to the (unwanted) **variability in system performance** and imply a look at how that may occur and which consequences it may have for the use of the system. In general, a highly reliable system is one which performs correctly according to specifications and where (unexplainable) performance variability consequently is low; conversely, a system with low reliability is one where performance variability is high. Considering a different meaning of the term, closer to the dictionary definition, the title may refer to the **trustworthiness** of the expert system (e.g. Muir, 1987) and the soundness (or validity) of the knowledge processing mechanisms (the rules) of it. And finally, the title may refer to the **quality** or **adequacy** of the expert system as a system component, e.g. the risks it may introduce, the safety with which it can be used, etc., cf. Chapter VI.

To simplify matters we shall just refer to the **internal** and **external** reliability of an expert system. The **internal reliability** is used to characterise the way in which the expert system works, i.e., the reliability of its internal functioning and mechanisms. This is exemplified by problems found in the representation of knowledge, the combination of numeric and semantic information, the handling of uncertainty and imprecision, and the nature of the inference engines. Applying the concepts introduced in Chapter I, the internal reliability corresponds to expert systems seen as part of the problem set.

In contrast to that the **external reliability** represents the view of the expert system that a user may have, i.e., how reliable the expert system is considered to be as a part of a solution. The typical problems here are how the expert system can be trusted (in the sense of being reliable), how it functions as a component in a larger context, and how the quality of its advices rates. The external reliability corresponds to expert

systems seen as part of the solution set.

The external reliability clearly depends on the internal reliability, although the relationship is far from straightforward. In most cases it is, of course, reasonable to assume that a high external reliability corresponds to a high internal reliability, and that a low external reliability corresponds to a low internal reliability. But while a high internal reliability is a necessary condition for a high external reliability it is not sufficient to guarantee it. A high reliability of the individual parts of the expert system does not necessarily produce a good result, because the way they interact may be defective. Users may furthermore distrust the expert system for other reasons, which have little or nothing to do with either kind of reliability. For instance, the interface between the system and the user or the human-computer interaction is clearly crucial for how well the expert system can be used. If it does not allow easy and flexible communication, the user may be unable to profit from the expert system's capabilities. Conversely, a low internal reliability does not necessarily result in a low external reliability. If, for instance, the expert system is new, i.e., the technology is in the growing phase, and the responses are sufficiently vague, the user may find it reliable because the enthusiasm for the expert system will colour the interpretation of the output and give the system the benefit of the doubt.

Finally, the performance and responses of the expert system should not only be judged in their own right, but should also be compared with the alternative, which in most cases is human advice and decisions. In many of the areas where expert systems are applied human decisions are quite often imprecise, but are generally followed by a mass of explanation that gives the impression of high reliability. A good example is expert advice on economic matters, e.g. investment, stock trade or currency exchange. For instance, a recent article about 'Investing in tomorrow's Europe' contained the following advice by a chief economist of a major European brokerage firm:

> *Private investors have to realise that the benefit of a single European market are here and now. This is the time to buy special situation stocks and take advantage of the intense level of takeover activity.*

While this is undoubtedly true, it is so vague that it is of very little practical use. If such advice was given by an expert system, it would most likely be discounted and considered close to useless. But being given by a human, it can be bolstered by additional information and explanations (dodges), which may make it palatable. Furthermore, we normally assume that because the speaker is an established expert in the domain, there must be something of value in the statement. It is therefore being

seen as part of a much larger context and given a favourable interpretation, rather than being treated on its own merits.

The book will cover the two main aspects of reliability, internal and external reliability, and try to show how they are related. Each of the four following chapters will focus on a particular facet of reliability and treat it in some detail, while the comments at the end will review the main issues and summarise the (possible) solutions. This introduction will start the discussion by considering the main issues in the use of expert systems with regard to their safety, risk, and reliability.

2. THE DEVELOPMENT IN THE USE OF EXPERT SYSTEMS

Expert systems can be defined in a number of ways, ranging from the very formal to the very informal. Following are some typical examples:

- Expert systems are computer systems which have a clear separation between a body of domain knowledge (using declarative coding) and reasoning strategies, i.e., the inference engine (using procedural coding). This definition, which is used in Chapter IV, highlights the structure and components of an expert system.

- An expert system can be defined, very informally, as a computer system that achieves high levels of performance in task areas that otherwise require years of special education and training. Expert systems furthermore allow the sharing of expertise between multiple users. This definition concentrates on the practical benefits of expert systems, and is used in Chapter V.

- Expert systems are computer systems that perform on the level of a human expert in autonomous knowledge-intensive (reasoning) tasks, such as diagnosis, analysis, decision making, problem solving, planning, etc. This definition puts the focus on the significant functional aspects of expert systems, and highlights the feature that is responsible for their very name.

Each definition reflects a particular point of view, either the strictly technical (the first definition), the pragmatic, utilitarian (the second definition), or the applied scientific (the third definition). The definitions are therefore not mutually exclusive but rather complementary. In addition, an often mentioned feature of expert systems is that they can be used by people who do not have a detailed knowledge of computer science, because the expert systems have flexible human-computer dialogues and problem-oriented languages for knowledge representation (Kriz, 1987). This resistivity of the