CHAPTER I: INTRODUCTION

A. ARTIFICIAL INTELLIGENCE

ARTIFICIAL INTELLIGENCE (AI) is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior—understanding language, learning, reasoning, solving problems, and so on. Many believe that insights into the nature of the mind can be gained by studying the operation of such programs. Since the field first evolved in the mid-1950s, AI researchers have invented dozens of programming techniques that support some sort of intelligent behavior. The *Handbook of Artificial Intelligence* is an encyclopedia of the major developments of the field's first 25 years—programs, programming techniques, and the computational concepts used to describe them.

Whether or not they lead to a better understanding of the mind, there is every evidence that these developments will lead to a new, *intelligent technology* that may have dramatic effects on our society. Experimental AI systems have already generated interest and enthusiasm in industry and are being developed commercially. These experimental systems include programs that

1. solve some hard problems in chemistry, biology, geology, engineering, and medicine at human-expert levels of performance,
2. manipulate robotic devices to perform some useful, repetitive, sensory-motor tasks, and
3. answer questions posed in simple dialects of English (or French, Japanese, or any other natural language, as they are called).

There is every indication that useful AI programs will play an important part in the evolving role of computers in our lives—a role that has changed, in our lifetimes, from remote to commonplace and that, if current expectations about computing cost and power are correct, is likely to evolve further from useful to essential.

The *Handbook* is composed of short articles about AI concepts, techniques, and systems, grouped into chapters that correspond to the major subdivisions of the field. This first article of the *Handbook* discusses what we mean by "artificial intelligence," both in terms of the interests and methods of the people doing AI research and in terms of the kinds of intelligent programs they have studied. We hope that this brief introduction will elucidate not only the potentially dramatic impact of intelligent technology on society, but also the possibilities that AI research affords for new insights into the puzzle that human intelligence is. The other two articles in this introductory chapter are meant as study guides: One describes the *Handbook* itself—its organization, what is in it and what is not—while the second offers guides for finding relevant material in the literature.

The Origins of Artificial Intelligence

> Scientific fields emerge as the concerns of scientists congeal around various phenomena. Sciences are not defined, they are recognized.
> (Newell, 1973a, p. 1)

The intellectual currents of the times help direct scientists to the study of certain phenomena. For the evolution of AI, the two most important forces in the intellectual environment of the 1930s and 1940s were *mathematical logic,* which had been under rapid development since the end of the 19th century, and new ideas about *computation.* The logical systems of Frege, Whitehead and Russell, Tarski, and others showed that some aspects of reasoning could be formalized in a relatively simple framework. Mathematical logic has continued to be an active area of investigation in AI, in part because logico-deductive systems have been successfully implemented on computers. But even before there were computers, the mathematical formalization of logical reasoning shaped people's conception of the relation between computation and intelligence.

Ideas about the nature of computation, due to Church, Turing, and others, provided the link between the notion of formalization of reasoning and the computing machines about to be invented. What was essential in this work was the abstract conception of computation as *symbol processing.* The first computers were numerical calculators that appeared to embody little, if any, real intelligence. But before these machines were even designed, Church and Turing had seen that numbers were an inessential aspect of computation—they were just one way of interpreting the internal states of the machine. Turing, who has been called the Father of AI, not only invented a simple, universal, and nonnumerical model of computation, but also argued directly for the possibility that computational mechanisms could behave in a way that would be perceived as intelligent. Douglas Hofstadter's book *Gödel, Escher, Bach: an Eternal Golden Braid* (1979) gives a thorough and fascinating account of the development of these ideas about logic and computation and of their relation to AI.

As Allen Newell and Herbert Simon point out in the "Historical Addendum" to their classic work *Human Problem Solving* (1972), there were other strong intellectual currents from several directions that

converged in the middle of this century in the people who founded the science of Artificial Intelligence. The concepts of cybernetics and self-organizing systems of Wiener, McCulloch, and others focused on the macroscopic behavior of "locally simple" systems. The cyberneticists influenced many fields because their thinking spanned many fields, linking ideas about the workings of the nervous system with information theory and control theory, as well as with logic and computation. The ideas of the cyberneticists were part of the Zeitgeist, and in many cases they influenced the early workers in AI directly as their teachers.

What eventually connected these diverse ideas was, of course, the development of the computing machines themselves, guided by Babbage, Turing, von Neumann, and others. It was not long after the machines became available that people began to try to write programs to solve puzzles, play chess, and translate texts from one language to another—the first AI programs. What was it about computers that triggered the development of AI? Many ideas about computing relevant to AI emerged in the early designs—ideas about memories and processors, about systems and control, and about levels of languages and programs. But the single attribute of the new machines that brought about the emergence of a new science was their very *complexity,* encouraging the development of new and more direct ways of describing complex processes—in terms of complicated data structures and procedures with hundreds of different steps.

Computers, Complexity, and Intelligence

As Pamela McCorduck notes in her entertaining historical study of AI, *Machines Who Think* (1979), there has been a long-standing connection between the idea of complex mechanical devices and intelligence. Starting with the fabulously intricate clocks and mechanical automata of past centuries, people have made an intuitive link between the *complexity* of a machine's operation and some aspects of their own mental life. Over the last few centuries, new technologies have resulted in a dramatic increase in the complexity we can achieve in the things we build. Modern computers are orders of magnitude more complex than anything man has built before.

The first work on computers in this century focused on the kinds of numerical computations that had previously been performed collaboratively by teams of hundreds of clerks, organized so that each did one small subcalculation and passed his results on to the clerk at the next desk. Not long after the dramatic success demonstrated by the first digital computers with these elaborate calculations, people began to explore the possibility of more generally intelligent mechanical behavior—could machines play chess, prove theorems, or translate languages?

They could, but not very well. The computer performs its calculations following the step-by-step instructions it is given—the method must be specified *in complete detail.* Most computer scientists are concerned with designing new algorithms, new languages, and new machines for performing tasks like solving equations and alphabetizing lists—tasks that people perform with methods they can explicate. However, people cannot specify how they decide which move to make in a chess game or how they determine that two sentences "mean the same thing."

The realization that the detailed steps of almost all intelligent human activity were unknown marked the beginning of Artificial Intelligence as a separate part of computer science. AI researchers investigate different kinds of computation and different ways of describing computation in an effort not just to create intelligent artifacts, but also to understand what intelligence is. Their basic tenet is that human intellectual capacity will be best described in the terms that we invent to describe AI programs. However, we are just beginning to learn enough about those programs to know how to describe them. The computational ideas discussed in this book have been used in programs that perform many different tasks, sometimes at the level of human performance, often much worse. Most of these methods are obviously not the same ones that people use to perform the tasks—some of them might be.

Consider, for example, computer programs that play chess. Current programs are quite proficient—the best experimental systems play at the human "expert" level, but not as well as human chess "masters." The programs work by searching through a space of possible moves, that is, considering the alternative moves and their consequences several steps ahead in the game, just as human players do. Computers can search through thousands of moves in the same time it takes human players to search through a dozen or so, and techniques for efficient searching constitute some of the core ideas of AI. The reason that computers cannot beat the best human players is that looking ahead is not all there is to chess—since there are too many possible moves to search exhaustively, alternatives must be evaluated without knowing for sure which will lead to a winning game, and this is one of those abilities that human experts cannot explicate. Psychological studies have shown that chess masters have learned to "see" thousands of meaningful configurations of pieces when they look at a chess position, which presumably helps them decide on the best move, but no one has yet designed a computer program that can identify these configurations.

The Status of Artificial Intelligence

Within most scientific disciplines there are several distinct areas of research, each with its own specific interests, research techniques, and terminology. In AI, these specializations include research on language understanding, vision systems, problem solving, AI tools and programming languages, automatic programming, and several others. As is apparent from its chapter headings, the *Handbook* is organized around the different subareas, as are most reviews of progress in AI. (See, e.g., Nilsson's thorough, and surprisingly current, survey of AI research, written in 1974.) The following discussion of the status of AI attempts to cut across the subfields, identifying some aspects of intelligent behavior and indicating the state of relevant AI research.

There is an important philosophical point here that we will sidestep. Doing arithmetic or learning the capitals of all the countries of the world, for example, are certainly activities that *indicate* intelligence in humans. The issue here is whether a computer system that can perform these tasks can be said to *know* or *understand* anything. This point has been discussed at length (see, e.g., Searle, 1980, and appended commentary), and we will avoid it here by describing the *behaviors* as intelligent and not concerning ourselves with how to describe the machines that produce them. Many intelligent activities besides numerical calculation and information retrieval have been accomplished by programs. Many key thought processes—like recognizing people's faces and reasoning by analogy—are still puzzles; they are performed so "unconsciously" by people that adequate computational mechanisms have not been postulated for them.

One word of caution. Like the different subfields of AI, the different behaviors discussed here are not at all independent. Separating them out is just a convenient way of indicating what current AI programs can do and what they can't do. Most AI research projects are concerned with many, if not all, of these aspects of intelligence.

Problem solving. The first big "successes" in AI were programs that could solve puzzles and play games like chess. Techniques like looking ahead several moves and dividing difficult problems into easier subproblems evolved into the fundamental AI techniques of *search* and *problem reduction.* Today's programs play championship-level checkers and backgammon, as well as very good chess. Another problem-solving program that integrates mathematical formulas symbolically has attained very high levels of performance and is being used by scientists and engineers across the country. Some programs can even improve their performance with experience.

As discussed above, the open questions in this area involve capabilities that human players have but cannot articulate, like the chess master's ability to see the board configuration in terms of meaningful patterns. Another basic open question involves the original conceptualization of a problem, called in AI the choice of problem representation. Humans often solve a problem by finding a way of thinking about it that makes the solution easy—AI programs, so far, must be told how to think about the problems they solve (i.e., the space in which to search for the solution).

Logical reasoning. Closely related to problem and puzzle solving was early work on logical deduction. Programs were developed that could "prove" assertions by manipulating a database of facts, each represented by discrete data structures just as they are represented by discrete formulas in mathematical logic. These methods, unlike many other AI techniques, could be shown to be complete and consistent. That is, so long as the original facts were correct, the programs could prove all theorems that followed from the facts, and only those theorems.

Logical reasoning has been one of the most persistently investigated subareas of AI research. Of particular interest are the problems of finding ways of focusing on only the relevant facts in a large database and of keeping track of the justifications for beliefs and updating them when new information arrives.

Language. The domain of language understanding was also investigated by early AI researchers and has consistently attracted interest. Programs have been written that answer questions posed in English from an internal database, that translate sentences from one language to another, that follow instructions given in English, and that acquire knowledge by reading textual material and building an internal database. Some programs have even achieved limited success in interpreting instructions spoken into a microphone instead of typed into the computer. Although these language systems are not nearly so good as people are at any of these tasks, they are adequate for some applications. Early successes with programs that answered simple queries and followed simple directions, and early failures at machine translation, have resulted in a sweeping change in the whole AI approach to language. The principal themes of current language-understanding research are the importance of vast amounts of general, commonsense *world knowledge* and the role of *expectations*, based on the subject matter and the conversational situation, in interpreting sentences.

The state of the art of practical language programs is represented by useful "front ends" to a variety of software systems. These programs accept input in some restricted form—they cannot handle some of the nuances of English grammar and are useful for interpreting sentences only within a relatively limited domain of discourse. There has been very limited success at translating AI results in language and speech understanding programs into ideas about the nature of human language processing.

Programming. Although perhaps not an obviously important aspect of human cognition, programming itself is an important area of research in AI. Work in this field, called *automatic programming,* has investigated systems that can write computer programs from a variety of descriptions of their purpose—examples of input/output pairs, high-level language descriptions, and even English descriptions of algorithms. Progress has been limited to a few, fully worked-out examples. Research on automatic programming may result not only in semiautomated software-development systems, but also in AI programs that learn (i.e., modify their behavior) by modifying their own code. Related work in the theory of programs is fundamental to all AI research.

Learning. Certainly one of the most salient and significant aspects of human intelligence is the ability to learn. This is a good example of cognitive behavior that is so poorly understood that very little progress has been made in achieving it in AI systems. There have been several interesting attempts, including programs that learn from examples, from their own performance, and from being told. But in general, learning is not noticeable in AI systems.

Expertise. Recently the area of expert systems, or "knowledge engineering," has emerged as a road to successful and useful applications of AI techniques. Typically, the user interacts with an expert system in a "consultation dialogue," just as he would interact with a human who had some type of expertise—explaining his problem, performing suggested tests, and asking questions about proposed solutions. Current experimental systems have achieved high levels of performance in consultation tasks like chemical and geological data analysis, computer system configuration, structural engineering, and even medical diagnosis. Expert systems can be viewed as intermediaries between human experts, who interact with the systems in "knowledge acquisition" mode, and human users who interact with the systems in "consultation mode." Furthermore, much research in this area of AI has focused on endowing these systems with the ability to *explain* their reasoning, both to make the consultation more acceptable to the user and to help the human expert find errors in the system's reasoning when they occur.

Current research deals with a variety of problems in the design of expert systems. These systems are built through the painstaking interaction of a domain expert, who may not be able to articulate all of his knowledge, and the systems designer: the knowledge-acquisition process is the big bottleneck in the construction of expert systems. Current systems are limited in scope and do not have the same sense as humans have about knowing when they might be wrong. New research involves using the systems to teach novices as well as to consult with practitioners.

Robotics and vision. Another part of AI research that is receiving increasing attention involves programs that manipulate robot devices. Research in this field has looked at everything from the optimal movement of robot arms to methods of planning a sequence of actions to achieve a robot's goals. Although more complex systems have been built, the thousands of robots that are being used today in industrial applications are simple devices that have been programmed to perform some repetitive task. Most industrial robots are "blind," but some see through a TV camera that transmits an array of information back to the computer. Processing visual information is another very active, and very difficult, area of AI research. Programs have been developed that can recognize objects and shadows in visual scenes, and even identify small changes from one picture to the next, for example, for aerial reconnaissance.

Systems and languages. In addition to work directly aimed at achieving intelligence, the development of new tools has always been an important aspect of AI research. Some of the most important contributions of AI to the world of computing have been in the form of spin-offs. Computer-systems ideas like time-sharing, list processing, and interactive debugging were developed in the AI research environment. Specialized programming languages and systems, with features designed to facilitate deduction, robot manipulation, cognitive modeling, and so on, have often been rich sources of new ideas. Most recently, several knowledge-representation languages—computer languages for encoding knowledge and reasoning methods as data structures and procedures—have been developed in the last five years to explore a variety of ideas about how to build reasoning programs. Terry Winograd's article "Beyond Programming Languages" (1979) discusses some of his ideas about the future of computing, inspired, in part, by his AI research.

Invitation

There has been much activity and progress in the 25-year history of AI. And there is more activity now than ever. AI is a relatively well-funded discipline, principally, in the United States, by the Defense Department's Advanced Research Projects Agency and other government agencies. There are active AI research groups in other countries, including Japan, Canada, Britain, France, Germany, Australia, Italy, and the USSR. Increasing research support is coming from the private sector, where interest in using and marketing AI programs is on the rise. The real shortage is people—there are only a few AI research groups in universities and corporate laboratories; in terms of the number of people involved, the field is still quite small.

So, let us end this introduction with an invitation to those of you who are not working in AI: to join us in this rapidly moving field. The excitement of creating a powerful new technology is coupled in AI, as perhaps in no other field these days, with the potential for stumbling upon new insights into one of the *big* questions: Physicists ask what kind of place this universe is and seek to characterize its behavior systematically. Biologists ask what it means for a physical system to be *living*. We in AI wonder what kind of information-processing system can ask such questions.