Chapter XVIII: **Artificial Intelligence: Retrospects** (Hofstadter, D. (1979). *Godel, Escher, Bach*. New York: Basic Books.

IN 1950, ALAN TURING wrote a most prophetic and provocative article on Artificial Intelligence. It was entitled "Computing Machinery and Intelligence" and appeared in the journal Mind. I will say some things about that article but I would like to precede t with some remarks about Turing the man.

Alan Mathison Turing born in London in 1912. He was a child full of curiosity and humor. Gifted in mathematics, he went Cambridge where his interests in machinery and mathematical logic cross-fertilized and resulted in his famous paper on "computable numbers", in which he invented the theory of Turing machines and demonstrated unsolvability of the halting problem; it was published in 1937. In the 1940's, his interests turned from the theory of computing machines to the actual building of real computers. He was a major figure in the development computers in Britain, and a staunch defender of Artificial Intelligence when it first came under attack. One of his best friends was David Champernowne (who later worked on computer composition of music). Champernowne and Turing were both avid chess players and invented "round-the-house" chess: after your move, run around the house-if you get back before your opponent has moved, you're entitled to another move. More seriously, Turing and Champernowne invented the first chess playing program, called "Turochamp". Turing died young, at 41--apparently of an accident with chemicals. Or some say suicide. His mother, Sara Turing, wrote his biography. From the people she quotes, one gets the sense that Turing was highly unconventional, even gauche in some ways, but so honest and decent that he was vulnerable to the world. He loved games, chess, children, and bike riding; he was a strong long-distance runner. As a student at Cambridge, he bought himself a second-hand violin and taught himself to play. Though not very musical, he derived a great deal of enjoyment from it. He was somewhat eccentric, given to great bursts of energy in the oddest directions. One area he explored was the problem of morphogenesis in biology. According to his mother, Turing "had a particular fondness for the Pickwick Papers", but "poetry, with the exception of Shakespeare's, meant nothing to him." Alan Turing was one of the true pioneers in the field of computer science.

The Turing Test

Turing's article begins with the sentence: "I propose to consider the question `Can machines think?'" Since, as he points out, these are loaded terms, it is obvious that we should search for an operational way to approach the question. This, he suggests, is contained in what he calls the "imitation game"; it is nowadays known as the Turing test. Turing introduces it as follows:

It is played with three people: a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A". The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try to cause C to make the wrong identification. His answer might therefore be

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the questions and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as “I am the woman, don’t listen to him!” to her answers, but it will avail nothing as the man can make similar remarks.
We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

After having spelled out the nature of his test, Turing goes on to mal some commentaries on it, which, given the year he was writing in, are qui sophisticated. To begin with, he gives a short hypothetical dialogue between interrogator and interrogatee:

Q: Please write me a sonnet on the subject of the Forth Bridge [a bridge over the Firth of Forth, in Scotland].
A: Count me out on this one. I never could write poetry. Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621. Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at RI. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.

Few readers notice that in the arithmetic problem, not only is there an inordinately long delay, but moreover, the answer given is wrong! This would be easy to account for if the respondent were a human: a mere calculational error. But if the respondent were a machine, a variety o explanations are possible. Here are some:

(1) a run-time error on the hardware level (i.e., an irreproducible fluke);
(2) an unintentional hardware (or programming) error which (reproducibly) causes arithmetical mistakes;
(3) a ploy deliberately inserted by the machine's programmer (or builder) to introduce occasional arithmetical mistakes, so as to trick interrogators;
(4) an unanticipated epiphenomenon: the program has a hard time thinking abstractly, and simply made "an honest mistake", which it might not make the next time around;
(5) a joke on the part of the machine itself, deliberately teasing its interrogator.

Reflection on what Turing might have meant by this subtle touch opens up just about all the major philosophical issues connected with Artificial Intelligence.

Turing goes on to point out that

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. . . . We do not wish to penalize the machine for its inability to shine in beauty competitions, nor to penalize a man for losing in a race against an airplane.

One of the pleasures of the article is to see how far Turing traced out each line of thought, usually turning up a seeming contradiction at some stage and, by refining his concepts, resolving it at a deeper level of analysis. Because of this depth of penetration into the issues, the article still shines after nearly thirty years of tremendous progress in computer development and intensive work in AI. In the following short excerpt you can see some of this rich back-and-forth working of ideas:

The game may perhaps be criticized on the ground that the odds are weighted too heavily against the machine. If the man were to try to pretend to be the machine he would clearly make a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.

Once the test has been proposed and discussed, Turing remarks:

The original question "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless, I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

Turing Anticipates Objections

Aware of the storm of opposition that would undoubtedly greet this opinion, he then proceeds to pick apart, concisely and with wry humor, a series of objections to the notion that machines could think. Below I list the nine types of objections he counters, using his own descriptions of them. Unfortunately there is not space to reproduce the humorous and ingenious responses he formulated. You may enjoy pondering the objections yourself, and figuring out your own responses.

(1) *The Theological Objection*. Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine tan think.
(2) *The "Heads in the Sand" Objection*. The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so.
(3) *The Mathematical Objection*. [This is essentially the Lucas argument.]
(4) *The Argument from Consciousness*. “Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants." [A quote from a certain Professor Jefferson.]

Turing is quite concerned that he should answer this serious objection full detail. Accordingly, he devotes quite a bit of space to his answer, and it he offers another short hypothetical dialogue:

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day", would not "a spring day" do as well or better?
Witness: It wouldn't scan.
Interrogator: How about "a winter's day"? That would scan all right.
Witness: Yes, but nobody wants to be compared to a winter's day.
Interrogator: Would you say Mr. Pickwick reminded you of Christmas?
Witness: In a way.
Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

After this dialogue, Turing asks, "What would Professor Jefferson say the sonnet-writing machine was able to answer like this in the *viva voce*?

Further objections:
(5) *Arguments from Various Disabilities*. These arguments take the form, "I grant you that you can make machines do all the things that you have mentioned but you will never be able to make one to do X." Numerous features X are suggested in this connection. I offer a selection:
Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make someone fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new.
(6) *Lady Lovelace's Objection*. Our most detailed information of Babbage's Analytical Engine comes from a memoir by Lady Lovelace. In it she states, "The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it to perform*" (her italics).
(7) *Argument from Continuity in the Nervous System*. The nervous system is certainly not a discrete state machine. A small error in the information about the size of a nervous impulse impinging on a neuron may make a large difference to the size of the outgoing impulse. It may be argued that, this being so, one cannot expect to be able to mimic the behaviour of the nervous system with a discrete state system.

(8) *The Argument from Informality of Behaviour*. It seems to run something like this. "If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines."

(9) *The Argument from Extra-Sensory Perception*. Let us play the imitation game, using as witnesses a man who is good as a telepathic receiver, and a digital computer. The interrogator can ask such questions as "What suit does the card in my right hand belong to?" The man by telepathy or clairvoyance gives the right answer 130 times out of 400 cards. The machine can only guess at random, and perhaps get 104 right, so the interrogator makes the right identification.

As you can see, many of the objections overlap and are related to each other. I have attempted in this book to answer most of them in some way or other, usually far less concisely than Turing did.

Objection (9) I find remarkable. I have seen the Turing paper reprinted in a book-but with objection (9) omitted-which I find equally remarkable. My own point of view-contrary to Turing's-is that ESP does not exist. Turing was reluctant to accept the idea that ESP is real, but did so nonetheless, being compelled by his outstanding scientific integrity to accept the consequences of what he viewed as powerful statistical evidence in favor of ESP. I disagree, though I consider it an exceedingly complex and fascinating question.

"Parry Encounters the Doctor"

In the Dialogue preceding this Chapter, you have seen an authentic exchange between a computer program and a human. Now there have been many programs with less ability to "converse" than SHRDLU. Two rather famous ones are "Doctor", created by Joseph Weizenbaum, and "Parry", created by Kenneth Colby. The former is supposed to simulate a psychiatrist using "nondirective therapy", the latter to simulate the belief structure of a paranoid. An amusing head-on collision between the two programs was arranged over a computer network, and the following is a short excerpt from the full transcript:

Parry: I don't understand your motives.
Doctor: Why don't you understand my motives?
Parry: Shouldn't I be?
Doctor: Why do you say that just now? Parry:
That's just the way it is.
Doctor: Suppose that's just the way it were not.
Parry: So?
Doctor: I am not sure I understand you fully.
Parry: You harp on the same questions too much.
Doctor: What does that similarity suggest to you? Parry: I don't get you.
Doctor: Do you wish that you did get I?
Parry: It's about time you showed some interest in my feelings.
Doctor: Oh? I showed some interest in your feelings?

These programs, of course, were made to talk to humans, so they are shown up rather poorly here, pitted against each other. Mostly they rely on shrewd guesses as to the nature of the input (which they analyze quite shallowly) and spit back canned answers that have been carefully selected from a large repertoire. The answer may be only partially canned: for example, a template with blanks that can be filled in. It is assumed that their human partners will read much more into what they say than is actually underlying it. And in fact, according to Weizenbaum, in his book *Computer Power and Human Reason*, just that happens. He writes:

ELIZA [the program from which Doctor was made] created the most remarkable illusion of having understood in the minds of the many people who conversed with it.... They would often demand to be permitted to converse with the system in private, and would, after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them.

Given the above excerpt, you may find this incredible. Incredible, but true Weizenbaum has an explanation:

Most men don't understand computers to even the slightest degree. So, unless they are capable of very great skepticism (the kind we bring to bear while watching a stage magician), they can explain the computer's intellectual feats only by bringing to bear the single analogy available to them, that is, their model of their own capacity to think. No wonder, then, that they overshoot the mark; it is truly impossible to imagine a human who could imitate ELIZA, for example, but for whom ELIZA's language abilities were his limit.

Which amounts to an admission that this kind of program is based on a shrewd mixture of bravado and bluffing, taking advantage of people's gullibility.

In light of this weird "ELIZA-effect", some people have suggested that the Turing test needs revision, since people can apparently be fooled by simplistic gimmickry. It has been suggested that the interrogator should be a Nobel Prize-winning scientist. It might be more advisable to turn the Turing test on its head, and insist that the interrogator should be another computer. Or perhaps there should be two interrogators-a human and Z computer-and one witness, and the two interrogators should try to figure out whether the witness is a human or a computer.

In a more serious vein, I personally feel that the Turing test, a: originally proposed, is quite reasonable. As for the people who Weizenbaum claims were sucked in by ELIZA, they were not urged to be skeptical; or to use all their wits in trying to determine if the "person" typing to them were human or not. I think that Turing's insight into this issue was sound, and that the Turing test, essentially unmodified, will survive.

A Brief History of AI
I would like in the next few pages to present the story, perhaps from an unorthodox point of view, of some of the efforts at unraveling the algorithms behind intelligence; there have been failures and setbacks and there will continue to be. Nonetheless, we are learning a great deal, and it i~ an exciting period.
Ever since Pascal and Leibniz, people have dreamt of machines that could perform intellectual tasks. In the nineteenth century, Boole and De Morgan devised "laws of thought" – essentially the Propositional Calculus-and thus took the first step towards AI software; also Charles Babbage designed the first "calculating engine"-the precursor to the hardware of computers and hence of AI. One could define AI as coming into existence at the moment when mechanical devices took over any tasks previously performable only by human minds. It is hard to look back and imagine the feelings of those who first saw toothed wheels performing additions and multiplications of large numbers. Perhaps they experienced a sense of awe at seeing "thoughts" flow in their very physical hardware. In any case, we do know that nearly a century later, when the first electronic computers were constructed, their inventors did experience an awesome and mystical sense of being in the presence of another kind of "thinking being". To what extent real thought was taking place was a source of much puzzlement; and even now, several decades later, the question remains a great source of stimulation and vitriolics.

It is interesting that nowadays, practically no one feels that sense of awe any longer-even when computers perform operations that are incredibly more sophisticated than those which sent thrills down spines in the early days. The once-exciting phrase "Giant Electronic Brain" remains only as a sort of "camp" cliche, a ridiculous vestige of the era of Flash Gordon and Buck Rogers. It is a bit sad that we become blase so quickly.

There is a related "Theorem" about progress in AI: once some mental function is programmed, people soon cease to consider it as an essential ingredient of "real thinking". The ineluctable core of intelligence is always in that next thing which hasn't yet been programmed. This "Theorem" was first proposed to me by Larry Tesler, so I call it *Tesler's Theorem*: "AI is whatever hasn't been done yet."

A selective overview of AI is furnished below. It shows several domains in which workers have concentrated their efforts, each one seeming in its own way to require the quintessence of intelligence. With some of the domains I have included a breakdown according to methods employed, or more specific areas of concentration.

mechanical translation
    direct (dictionary look-up with some word rearrangement)
        indirect (via some intermediary internal language)
game playing
    chess
        with brute force look-ahead
        with heuristically pruned look-ahead with no look-ahead
    checkers
    go
    kalah
    bridge (bidding; playing)
    poker
    variations on tic-tac-toe
    etc.

proving theorems in various parts of mathematics
    symbolic logic
        "resolution" theorem-proving
    elementary geometry
symbolic manipulation of mathematical expressions
    symbolic integration
    algebraic simplification
    summation of infinite series
vision
    printed matter:
        recognition of individual hand-printed characters drawn from a small class (e.g., numerals)
        reading text in variable fonts
        reading passages in handwriting
        reading Chinese or Japanese printed characters
        reading Chinese or Japanese handwritten characters
pictorial
    locating prespecified objects in photographs
    decomposition of a scene into separate objects
    identification of separate objects in a scene
    recognition of objects portrayed in sketches by people
    recognition of human faces
hearing
    understanding spoken words drawn from a limited vocabulary (e.g., names of the ten digits)
    understanding continuous speech in fixed domains
    finding boundaries between phonemes
    identifying phonemes
    finding boundaries between morphemes
    identifying morphemes
    putting together whole words and sentences
understanding natural languages
    answering questions in specific domains
    parsing complex sentences
    making paraphrases of longer pieces of text using knowledge of the real world in order to understand passages
    resolving ambiguous references

producing natural language
    abstract poetry (e.g., haiku)
    random sentences, paragraphs, or longer pieces of text
    producing output from internal representation of knowledge

creating original thoughts or works of art
- poetry writing (haiku)
- story writing
- computer art
- musical composition
  - atonal
  - tonal

analogical thinking
- geometrical shapes ("intelligence tests")
- constructing proofs in one domain of mathematics based on those in a related domain

learning
- adjustment of parameters
- concept formation

Mechanical Translation

Many of the preceding topics will not be touched upon in my selective discussion below, but the list would not be accurate without them. The first few topics are listed in historical order. In each of them, early efforts fell short of expectations. For example, the pitfalls in mechanical translation came as a great surprise to many who had thought it was a nearly straightforward task, whose perfection, to be sure, would be arduous, but whose basic implementation should be easy. As it turns out, translation is far more complex than mere dictionary look-up and word rearranging. Nor is the difficulty caused by a lack of knowledge of idiomatic phrases. The fact is that translation involves having a mental model of the world being discussed, and manipulating symbols in that model. A program which makes no use of a model of the world as it reads the passage will soon get hopelessly bogged down in ambiguities and multiple meanings. Even people-who have a huge advantage over computers, for they come fully equipped with an understanding of the world-when given a piece of text and a dictionary of a language they do not know, find it next to impossible to translate the text into their own language. Thus-and it is not surprising in retrospect-the first problem of AI, led immediately to the issues at the heart of AI.