Turing Test and After

V Rajaraman

V Rajaraman is with the Jawaharlal Nehru Centre for Advanced Scientific Research, and the Indian Institute of Science, Bangalore. Several generations of scientists and engineers in India have learnt to program using his lucidly written books on programming.

In this article we discuss a test described by A M Turing to find out whether a computer program exhibits intelligence and also of its modified versions used to evaluate expert systems and natural language conversations generated by computer programs.

Alan Mathison Turing was conceived in India (where his father Julius Mathison Turing was an officer in the Indian Civil Service) and was born on June 23, 1912 in Paddington, U.K. He was educated entirely in England and got his degree in mathematics from Cambridge University in 1934. His scientific career was marked by three landmark contributions. The first in 1935 was on computable numbers, submitted to London Mathematical Society on May 28, 1936. The second was in 1940 during the second world war when he found a method of breaking the German enigma code (see Whitemore in Suggested Reading) and designed a machine to do it. The last, before his premature death in 1954, was his paper entitled 'Computing Machinery and Intelligence' which appeared in the Philosophical Magazine, *Mind*, in October 1950 (see Turing in Suggested Reading). Turing had been thinking about the 'enigma of intelligence' since his seminal work on the universal Turing Machine in 1935 and his subsequent attempts to build a stored program computer (called Automatic Computing Engine (ACE)) at the National Physical Laboratories in U.K. during 1947.

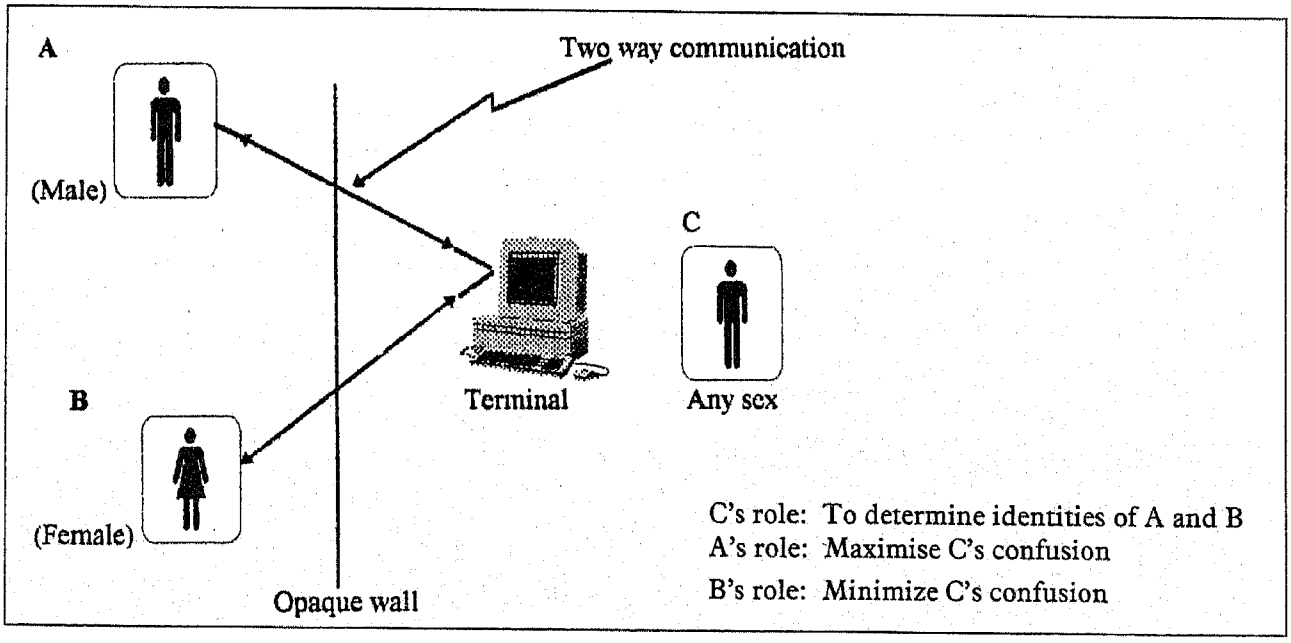
Can Machines Think?

While working at the University of Manchester in 1949 where he had gone from National Physical Laboratory (as there was a working computer of reasonable size there) he was convinced that machines could be programmed to exhibit intelligent

behaviour. In support he wrote an article in *Mind* (see Turing in Suggested Reading). The article was written in a conversational style just as he would have argued the topic with a friend. The article starts with the sentences: "I propose to consider the question, "Can machines think?" This should begin with definitions of the meanings of the terms "machines" and "think"."

He quickly concludes that the common usage of these words are vague and we cannot answer the question in its current form. He thus replaced the question by another less ambiguous one and called it the 'imitation game'. This game is to be played by three human beings – a man labelled A, a woman labelled B and an interrogator labelled C. The interrogator C cannot see A or B but can pose questions to A or B typing them on a teleprinter. From the answers, C is to decide which is the man and which is the woman. In the game the man will respond like a woman to fool C but the woman will respond truthfully to convince C that she is indeed a woman. He proposed a variation of this game in which a computer program acts as A. The interrogator does not know this and poses his questions using a teleprinter to A and B and tries to determine which respondent is the computer (see *Figure 1*). Unfortunately this part of the paper is vague and Turing does not state clearly the details of the game such as how

Figure 1 Turing's imitation game.



The Turing test provides an unbiased method of comparing intelligent behaviour of humans with that of computers.

many questions are allowed, when does the game end, does A know the questions posed to B and her answers.

Turing Test

A modified version of the imitation game known as *Turing Test* will be described now. The interrogator is in a room (see *Figure 2*) and poses questions using a video terminal to a computer and a human. The interrogator cannot see who types the answers which appear on his terminal.

If after some sessions the interrogator is unable to decide which answers were given by the computer and which by the human, then we can conclude that the human and the machine are indistinguishable. In other words a machine exhibits 'intelligence' just like a human being.

The Turing test provides an unbiased method of comparing intelligent behaviour of humans with that of computers. The Turing test is repeatable and objective. Turing did not regard his test as a necessary condition for attributing intelligence to machines but only as a sufficient condition. He stated his belief that "in about fifty years time it will be possible to program computers with a storage capacity about 10^9, to make them play the imitation game so well that an average interrogator will not

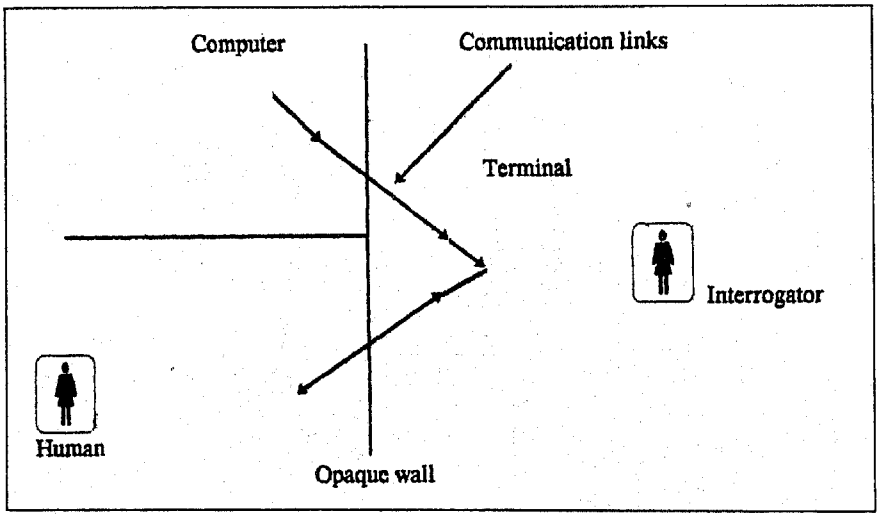


Figure 2 Turing test set up.

have more than 70 percent chance of making the right identification after five minutes of questioning".

Turing was aware that even though his original question had been replaced by a specific test there may still be doubts whether his test is a reasonable replacement of the original problem. He then considered various objections one may raise to this, ranging from theological objections to objections based on extrasensory perception. In one of his arguments he imagines a conversation (see *Box* 1) between an interrogator and a witness. If the witness were a computer what would the interrogator conclude about its abilities?

In a restricted Turing test the topic of the conversation is restricted and one is expected to write a program whose answers to queries on the topic would be most human like.

Restricted Turing Test

As of today no computer can answer like the witness (*Box 1*) as it would require an enormous knowledge base and 'common sense'. Thus most of the work has been to program a computer to pass a restricted Turing test. In a restricted Turing test the topic of the conversation is restricted to a narrow subject such as jazz music, cricket personalities and scores etc., and one is expected to write a program whose answers to queries on the topic would be 'most human like'. One of the requirements of Turing test is

Box 1 Part of an oral examination imagined by Turing.

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 are serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

Box 2 Sample of ELIZA's conversation.

(Human's statements: Upper and lower case. Computer's statements: Upper case)
Men are all alike
IN WHAT WAY
They're always bugging us about something or other
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOY FRIEND MADE YOU COME HERE.
He says I'm depressed much of the time
I AM SORRY TO HEAR YOU ARE DEPRESSED
It's true I am unhappy
DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY
I need some help, that much seems certain
WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP
Perhaps I could learn to get along with my mother
TELL ME MORE ABOUT YOUR FAMILY
My mother takes care of me
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU
My father
YOUR FATHER
---

the ability to converse in a natural language. An early attempt at natural language conversation by a computer was a program called ELIZA written in 1966 by Joseph Weizenbaum (see Suggested Reading). ELIZA was to mimic a non directive psychotherapist. In *Box 2*, a typical ELIZA conversation is given. ELIZA was developed using a very simple principle. There is a small set of keywords which appear in most conversations. ELIZA keeps sentence templates which are used to print a sentence whenever one such keyword appears. For example, a reference to a member of one's family such as "my brother" or "my wife" leads to printing of the sentence "Tell me more about your family". If "my" refers to anything other than a family member ELIZA stores the word string following "my" and interchanges first and second person pronouns and

One of the requirements of Turing test is the ability to converse in a natural language.

Box 3 Can Deep Blue Pass Chess Restricted Turing Test?

Fredric Friedel , Kasparov's computer adviser, conducted an informal experiment and showed Kasparov a series of games in a tournament played by Deep Thought (the predecessor of Deep Blue, the current champion chess playing computer) and several grandmasters. Without identifying the players, Friedel asked Kasparov to pick out the moves made by the computer. In a number of cases Kasparov mistook the computer's moves for those of a grandmaster, or vice versa. In general, only chess players who have considerable experience playing against computers can identify computer moves. (Quoted from M S Campbell's book. See Suggested Reading.)

possessives in this string. Thus if at any point the interrogator types "It disturbs me that my hair is becoming grey", ELIZA might say, after many lines of conversation, "Does that have anything to do with the fact that your hair is becoming grey?" Even though ELIZA is quite shallow, it is able to easily fool many ordinary people into thinking that a person is answering the questions. Weizenbaum wrote in his book (see Suggested Reading) that he was startled to see how people conversing with his program became emotionally involved with it. Even his secretary, who had watched him work with the program, wanted him to leave the room while she conversed with the computer!

The term *restricted Turing test* is nowadays used to evaluate human-like behaviour of a variety of computer applications where artificial intelligence techniques are used (see Moor in Suggested Reading). For example, programs have been written to play chess. (In fact Turing wrote one of the very early chess playing programs). To evaluate how well such a program plays chess one may apply a 'chess restricted Turing test' to the program. In such a test a human judge is given the moves made by a computer and those made by a human in a game. If the judge is not able to distinguish between the two, one may say that the chess playing program has passed a chess restricted Turing test (see *Box* 3).

The term *restricted Turing test* is nowadays used to evaluate human-like behaviour of a variety of computer applications where artificial intelligence techniques are used.

Restricted Turing test has been used to probe the strengths and weaknesses of computer models imitating human behaviour.

Restricted Turing test has been useful in evaluating expert systems, i.e. programs which mimic the behaviour of an expert. In this context judges are asked to compare the behaviour of an expert system with that of a human expert and rank the system. For example, an expert system called MYCIN (see Rolston in Suggested Reading) was designed to diagnose and recommend treatment for meningitis. A panel of doctors was requested to evaluate the performance of MYCIN on a series of cases with that of physicians of various levels of training and experience. In this test MYCIN's diagnosis and treatment was rated to be equivalent or better than actual physicians.

Restricted Turing test has been used to probe the strengths and weaknesses of computer models imitating human behaviour. For example, the responses of PARRY (see Colby and others in Suggested Reading), a program that simulates paranoid behaviour, were compared with that of an actual patient by a panel of psychiatrists. PARRY was identified as the real paranoid patient by about half of them.

It may be observed from the above discussions that researchers in artificial intelligence have had reasonable success in developing systems which exhibit intelligent behaviour in restricted domains. It has, however, not been possible to create a system which integrates knowledge from diverse fields and incorporates 'commonsense knowledge' to be able to pass an unrestricted Turing test. Turing's conjecture that the test will be passed by the year 2000 seems over optimistic now. Some real breakthroughs are required before this happens and I hazard a guess that it may not happen for another 20 years.

Loebner Prize

In 1991, Hugh Loebner (a New York theatre equipment manufacturer) established a grand prize of $100,000 for a computer program that could pass an unrestricted Turing test.

As it was realised that no program would pass such a test as of now a reduced prize of $2000 and a bronze medal was announced for a computer program which passes a restricted Turing test (see Shieber in Suggested Reading). The prize is to be awarded every year to that program which is rated nearly human by a panel of non-specialist judges. The prize committee spent almost two years in planning the structure of the competition. There will be a number of contestants (computer programs), several confederates (human beings against whom the computer programs are judged) and judges. The judges and confederates are chosen from among lay men with no special expertise in computer science. The ground rules specified that the topic of the conversation should be strictly limited for both contestants and confederates. Judges should stay in the topic of the conversation and respond naturally without resorting to any 'trickery or guile'.

It has not been possible to create a system which integrates knowledge from diverse fields and incorporates 'commonsense knowledge' to be able to pass an unrestricted Turing test.

The committee has devised an ingenious scoring mechanism to adjudge the winner. Each human judge is required to place all the agents with whom he or she conversed in the order from the 'least human like' to 'most human like'. The contestant with the highest average rank is deemed the winner of the contest (*Figure* 3).

The judges are also requested to draw a vertical line separating those agents they consider as computer programs and those they consider as humans. In *Figure* 3 terminals H and E are identified by a judge as being operated by humans. In the first competition there were 6 contestants (computer programs) who

Figure 3 Form used by judges to score the agents in Loebner Prize.

Least human like							Most human like	
Rank	1	2	3	4	5	6	7	8
Terminal identity	B	C	A	D	F	G	H	E

There is no chance of anyone writing a program (at present) which would pass the unrestricted Turing test.

picked their topic of conversation as burgundy wines, dry martinis, small talk, whimsical conversation, child's birthday party and dissatisfaction in relationships. The confederates (humans) chose to converse on Shakespeare and women's clothing. At the end of the contest the average ranking identified the two human confederates as most human like. Joseph Weintraub's program (whimsical conversations) was awarded the prize.

Sheiber wrote a critique of the Loebner Prize contest in which he maintained that giving a restricted Turing test is not a good method of advancing research in Artificial Intelligence. He pointed out that Weintraub exploited a loophole in the competition rules by his choice of topic, namely, whimsical conversation. Schieber states that "when Weintraub's program is unresponsive, fails to make any sense, or shows a reckless abandonment of linguistic appropriateness, it, unlike its competitor programs, is operating as advertised". Weintraub won with the topic 'men vs women' in 1992 and he won again in 1993 using the same technique. In response Loebner maintained that his prize was a beginning to promote understanding of machine intelligence and that in the future the prize will be given for passing unrestricted Turing test only. The competition is, however, being held each year with restricted test as everyone agrees that there is no chance of anyone writing a program (at present) which would pass the unrestricted Turing test. The next competition is to be held on January 11, 1998 in Sydney, Australia.

Conclusions

Turing felt that his test would pave the way towards creation of a computer with a general intelligence. In the concluding section of his paper (*Box* 4) he argues that the method to construct a machine to pass the Turing test was to build a machine with some knowledge base and 'educate' it, may be through a question-answer regimen - rewarding correct answers