First AI Debate : Is AI possible??

- AI is the art of making machines do things that would require intelligence if done by a human, such as playing chess, speaking English, or diagnosing an illness.
- Sloman (1986) gives a broader view : "AI is a very general investigation of the nature of intelligence and the principles and mechanisms required for understanding or replicating it."
- This view exhibits both the engineering side of AI and its cognitive science side.
- The first AI debate are intimately concerned with each side..

Is AI possible on Computers?

- *Can we, or can we not, expect computers to think in the sense that humans do?*

- AI proponents say that consciousness will *emerge* in sufficiently complex machines.

Roger Penrose

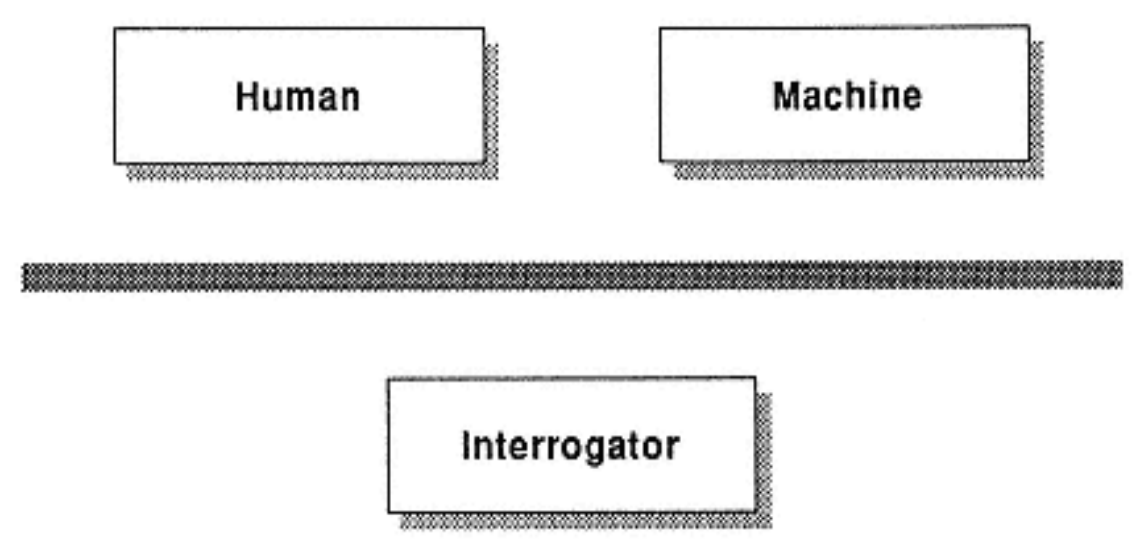
- *With thought comprising a non-computational element, computers can never do what we human beings can.* —Roger Penrose, *The Emperor's New Mind*
- Consciousness seems to me to be such an important phenomenon that I simply cannot believe that it is something just "accidentally" conjured up by a complicated computation. . . . . it is indeed "obvious" that the conscious mind cannot work like a computer, even though much of what is actually involved in mental activity might do so.
- Penrose is drawing the line at consciousness, claiming that a computer can never be conscious.

Moravec, 1988

- Today, our machines are still simple creations, requiring the parental care and hovering attention of any newborn, hardly worthy of the word "intelligent." But within the next century they will mature into entities as complex as ourselves, and eventually into something transcending everything we know—in whom we can take pride when they refer to themselves as our descendants. . . . We are very near to the time when virtually no essential human function, physical or mental, will lack an artificial counterpart. The embodiment of this convergence of cultural developments will be the intelligent robot, a machine that can think and act as a human.

The Turing Test

- Turing (1950) : a sufficient criterion for machine intelligence



- The machine will be deemed intelligent if sophisticated interrogators, unconstrained as to subject matter, cannot reliably tell which responder is human and which is machine.

Artificial mind via symbolic AI

- A physical symbol system uses physical symbols, symbolic structures (expressions) composed of these symbols, and operators that create, copy, modify, or destroy these symbols and/or structures. The system runs on a machine that, through time, produces an evolving collection of symbol structures.
- SOAR is one such physical symbol system.

Pro: Herbert Simon and Allen Newell

- "intelligence is the work of symbol systems" and "a physical symbol system has the necessary and sufficient means for general intelligent action. The computer is . . . [a] physical symbol system. . . . The most important [such] is the human mind and brain" (Simon 1981, p. 28).
- . . . there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until—in a visible future—the range of problems they can handle will be coextensive with the range to which the human mind has been applied. (Simon and Newell 1958)

The Dreyfus Attack

- AI has stumbled over commonsense knowledge and reasoning. A typical human, in our society, would expect a dropped glass of water to break and splatter its contents. A typical AI program would expect no such thing unless specifically instructed.
- Dreyfus (1988,p.33) "If background understanding is indeed a skill and if skills are based on whole patterns and not on rules, we would expect symbolic representations to fail to capture our commonsense understanding." Dreyfus (1988, p. 37) specifically attacks the physical symbol hypothesis: "The physical symbol system approach seems to be failing because it is simply false to assume that there must be a theory of every domain."

The Dreyfus Attack...

- A theory in this context is a collection of rules describing behavior within the domain. Dreyfus claims that not every domain is rule describable. Horgan and Tienson (1989, pp. 154ff.) offer basketball as such a domain. Imagine all the factors influencing a point guard's decision, during a fast break, to shoot or pass, and if the latter, to whom. They don't prove a description via rules to be impossible, but they clearly place the onus on proponents of rules to produce one.
- Decisions are not based on rules only.

Dreyfus identifies five stages of learning.

- From Novice to Expert : at each stage you add youe experience (say, to learn to drive a car) :
- (1) the *novice* uses rules, typically supplied by a teacher, applied to context-free features, also usually from the teacher;
- (2) a *beginner*, in addition, begins to recognize new situational aspects, such as using engine sounds to help determine when to shift gears;
- (3) a *competent* driver will examine only the situational features that are relevant to the selected goals or plan;
- (4) a *proficient* driver doesn't have to examine anymore but sees directly what is relevant—nonetheless, he or she decides consciously what to do at this point;
- (5) the *expert* just does it.

The Dreyfus Attack…

- . . . experience-based, holistic, similarity recognition produces the deep situational understanding of the proficient performer. No new insight is needed to explain the mental processes of the expert. (Dreyfus 1987, p. 102)
- Dreyfus maintains that similarity recognition says it all.
- Dreyfus believes that human experts typically choose the behavior that usually works via similarity recognition, without resorting to problem solving by means of rules.
- This directly contradicts the physical symbol system hypothesis of Simon and Newell by denying its necessity. It also calls into question its sufficiency, since humans provide the only currently known example of general intelligence

Searle's Chinese Room thought experiment

- First, Schank's notion of a script :

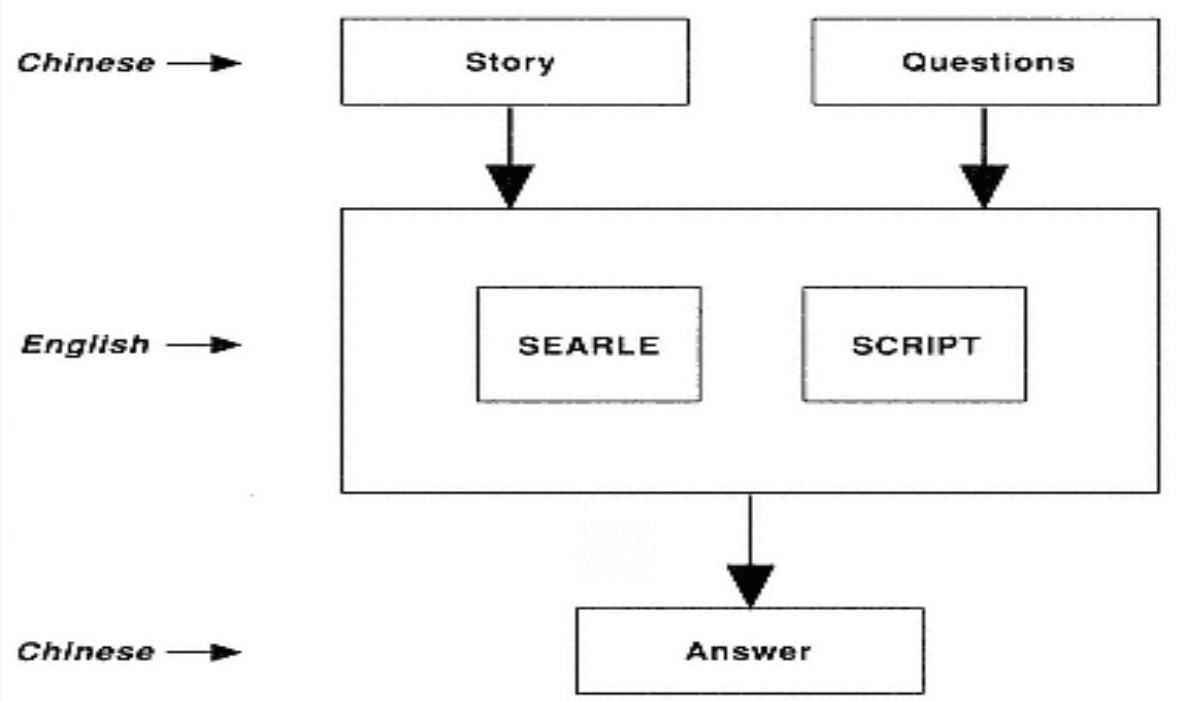
  Schank has been a leading exponent of natural language comprehension via machine. The data structure underlying several of his systems is called a *script*.

  Scripts are used for representing knowledge of common sequences of events, say those involved in going out to dinner. Such a script might record that you typically go into a restaurant, sit down (or are seated by a hostess), are brought a menu by a waiter, etc.

  Scripts may contain entry conditions, results, props, roles, tracks, scenes, and so on.

Searle's Chinese Room thought experiment…

- John Searle (1980), focuses on understanding



Searle's Chinese Room thought experiment…

- Searle puts himself in a closed room;
- A story and a question about it, both written in Chinese, are slipped under the door to him. Searle understands no Chinese;
- he has been given a comprehensive script, written in English, that provides an algorithmic way of answering the question
- Thinking in English, and carefully following the directions contained in the script, Searle produces squiggles that form a native speaker's answer to the question.
- Searle slips the results of this purely formal transformation under the door as an answer.
- Searle, who understands English but no Chinese, has produced an answer giving the appearance of understanding. He actually understands nothing of the story, the question, or the answer.
- In the same way, Searle asserts, Schank's computer understands nothing of the stories even as it answers questions about them.

Chinese Room Expt. : systems reply

- Searle doesn't understand, but the entire system, including the script, must understand because it produces native speaker answers. This leads to the question: What we mean by ‘Understanding’.
- How do we convince ourselves that some other person understands a story we've told?
- A common approach is to question that person and judge understanding by the quality of the answers.
- Searle counters the systems reply by supposing that he, Searle, memorizes the entire system. Then he is the system, and he still doesn't understand Chinese.

  Can Searle memorize that script?

Chinese Room Expt. : the brain simulator reply

- Suppose a computer program simulates the actual neural firings of a native Chinese speaker in order to answer questions posed about a simple story.
- To say that the system doesn't understand is to say that the native speaker doesn't understand, as the system is doing exactly what the native speaker did.
- Searle's counter : Suppose Searle calculates the answers using a system of water pipes and valves that simulate the native speaker's brain. Neither Searle, nor the water pipe system, nor their conjunction understands. The conjunction fails because the whole system could again be internalized. Searle could simply memorize all the workings of the native speaker's brain and go on from there. Thus, Searle maintains that the original script could have worked by simulating a native speaker's brain with the same lack of understanding.

Chinese Room Expt. …..

Many such Replies and many such counters….

Machines just don't understand

- Searle

Degrees of Understanding

- Arnie is a formal system (computer program) that solves quadratic equations. Feed it the numerical coefficients, a, b, and c of a quadratic equation $ax^2 + bx + c = 0$, and it pops out a solution.
- Arnie gives right answers but understands nothing of the numbers it spews, and nothing else of algebra.
- In Stan Franklin's view, it must understand something of quadratic equations because it can solve them correctly. It understands more than a program that produces incorrect solutions. But that's certainly very little understanding.
- Say for example in human beings there can be various levels of understanding of quadratic eqn., numbers and algebra.
- A system's understanding of a concept, or of a collection of concepts, seems to vary with the complexity of its connections from the given concepts to other knowledge. Roughly, the more connections, the more understanding

Degrees of Understanding…

- Searle anticipated this argument by degrees of understanding, and replied that computer systems have *zero* understanding, making them different in kind from the human systems with varying degrees of understanding.
- There must be at least as many levels of understanding Chinese, depending on the surrounding network of knowledge.
- It seems that the system comprised of Searle and his script does understand Chinese at some minimal level.
- Whether this artificial understanding must of necessity always fall short of that of humans, or whether the understanding of our "mind children" will eventually surpass ours, is the essence of this first AI debate (Stan Franklin)

Gödel's Incompleteness Theorem

- Context :

  Around the turn of the twentieth century, logicians, pushed by Hilbert (1901), were trying to derive mathematics from logic. The idea was to develop a formal system, starting with finitely many axioms and rules of deduction, and from it to deduce all of mathematics. One major attempt was produced by Frege (1893, 1903). Whitehead and Russell spent decades writing their version, the *Principia Mathematica* [12] (1910–1913). Then, in 1931, along came an unknown Austrian, Kurt Gödel, who proved that their whole endeavor was hopeless to begin with.

Gödel's Incompleteness Theorem

- Theorem :

  every sufficiently powerful formal theory allows a true but unprovable proposition (Gödel 1931)

  - "Sufficiently powerful" means that the system can account for integer arithmetic

  - a true but unprovable proposition P : A formal theory will have symbols (variables, constants, and operators) as well as rules of syntax. P must be composed of some of these symbols arranged according to the syntax of the system. In technical terms, it must be a well-formed formula of the system. Furthermore, P is unprovable from the axioms and rules of deduction of the system.[14] On the other hand, the meaning of P is known to be true in the system. P, for example, might assert that P is unprovable in the system.

Gödel's Incompleteness Theorem

- Gödel's theorem (computational form)

  every algorithm for deciding mathematical truth must fail to decide some proposition correctly
- An algorithm will tell you a proposition is true only if it can prove it. A true but unprovable proposition gets a wrong reading.
- The bottom line is that not all of mathematics can be produced algorithmically.

The Penrose Attack (Consciousness)

- (1) the nonalgorithmic nature of mathematical thought and (2) quantum mechanical effects in the brain. (relates to physics of quantum gravity)
- that brain processes involving frequent quantum decisions would be noncomputable. Hence, computers couldn't implement these processes, and artificial intelligence would be inherently limited
- All this is based on there being such quantum mechanical decisions in the brain, for which Penrose argues in detail, and on a nonalgorithmic theory of quantum gravity, which so far is pure speculation, though he provides plausibility arguments.

The Horgan-Tienson Attack

- Two philosophers, Horgan and Tienson, although not ruling out machine intelligence, claim that it cannot be produced via symbolic AI alone, via rule-based systems (1989).
- The rule-based systems cannot handle the following which are easily solved by mind :

  (a) multiple soft constraints, (b) cognitive folding and (c) the frame problem.
- Many activities requiring intelligence seem to involve the satisfaction of multiple soft constraints, e.g. going for shopping to a mall : arriving at the mall, searching out a parking place following prescribed traffic patterns, modified by the actions of other vehicles, and occupying it, making way to an entrance, taking care not to become a casualty on the way, etc. Surprisingly, most often people arrive at the mall without mishap.

The Horgan-Tienson Attack

- Adding constraints makes the task harder for computers but easier for humans. For computers, every trial solution must be tested against all applicable constraints. But to decide which constraints are applicable means, essentially, to test against all constraints. For humans, having more constraints sometimes makes things easier.
- The constraints, in addition to being multiple, are usually soft, means, each of the constraints has innumerable possible exceptions which requires other rules.

The Horgan-Tienson Attack

- Cognitive folding: Horgan and Tienson claim that cognition cannot be partitioned into isolated domains. Any part of commonsense knowledge might be called upon in dealing with any of a vast number of cognitive tasks (combining them).
- To pass the Turing test, a system must converse intelligently about many domains, individually and in every relevant combination. A sophisticated interrogator will see to it. But, say Horgan and Tienson, "there is little reason to believe that there are . . . domain independent rules[20] that govern the folding together of knowledge about any two . . . arbitrarily chosen domains" (1989, p. 153).

The Horgan-Tienson Attack

- The frame problem: to determine in an effective and general way, what to change and what to leave the same in a system of beliefs, when any new bit of information is added. Humans deal with this problem rather well; symbolic AI, with great difficulty and little success.
- E.g. a robot with a key in its hand. The system notes that the robot has passed from the office to the shop. Should it also note that the key is no longer located in the office? If so, how is it to do that? Building large systems that sensibly track such side effects seems an insurmountable problem.
- Could this be, in part, because we humans don't store representations of information but, rather, re-create those representations as needed? To a computer scientist, it's the difference between consulting a lookup table and running an algorithm

What do we conclude?

- Many weaknesses of Symbolic AI……