

Computing Machinery and Intelligence

"**Computing Machinery and Intelligence**" is a seminal paper written by Alan Turing on the topic of artificial intelligence. The paper, published in 1950 in Mind, was the first to introduce his concept of what is now known as the Turing test to the general public.

Turing's paper considers the question "Can machines think?" Since the words "think" and "machine" cannot be defined in a clear way that satisfies everyone, Turing suggests we "replace the question by another, which is closely related to it and is expressed in relatively unambiguous words."^[1] To do this, he must first find a simple and unambiguous idea to replace the word "think", second he must explain exactly which "machines" he is considering, and finally, armed with these tools, he formulates a new question, related to the first, that he believes he can answer in the affirmative.

강제, 응대

Contents

Turing's test

Digital machines

Nine common objections

Learning machines

Notes

References

External links

Turing's test

Rather than trying to determine if a machine is thinking, Turing suggests we should ask if the machine can win a game, called the "Imitation Game". The original Imitation game that Turing described is a simple party game involving three players. Player A is a man, player B is a woman and player C (who plays the role of the interrogator) can be of either sex. In the Imitation Game, player C is unable to see either player A or player B (and knows them only as X and Y), and can communicate with them only through written notes or any other form that does not give away any details about their gender. By asking questions of player A and player B, player C tries to determine which of the two is the man and which is the woman. Player A's role is to trick the interrogator into making the wrong decision, while player B attempts to assist the interrogator in making the right one.

Turing proposes a variation of this game that involves the computer: "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?'^[2] So the modified game becomes one that involves three participants in isolated rooms: a computer (which is being tested), a human, and a (human) judge. The human judge can converse with both the human and the computer by typing into a terminal. Both the computer and human try to convince the judge that they are the human. If the judge cannot consistently tell which is which, then the computer wins the game.^[3]

As Stevan Harnad notes,^[4] the question has become "Can machines do what we (as thinking entities) can do?" In other words, Turing is no longer asking whether a machine can "think"; he is asking whether a machine can act indistinguishably^[5] from the way a thinker acts. This question avoids the difficult philosophical problem of pre-defining the verb "to think" and focuses instead on the performance capacities that being able to think makes possible, and how a causal system can generate them.

Some have taken Turing's question to have been "Can a computer, communicating over a teleprinter, fool a person into believing it is human?"^[6] but it seems clear that Turing was not talking about fooling people but about generating human cognitive capacity.^[7]

Digital machines

Turing also notes that we need to determine which "machines" we wish to consider. He points out that a human clone, while man-made, would not provide a very interesting example. Turing suggested that we should focus on the capabilities of digital machinery—machines which manipulate the binary digits of 1 and 0, rewriting them into memory using simple rules. He gave two reasons.

First, there is no reason to speculate whether or not they can exist. They already did in 1950.

Second, digital machinery is "universal". Turing's research into the foundations of computation had proved that a digital computer can, in theory, simulate the behaviour of any other digital machine, given enough memory and time. (This is the essential insight of the Church–Turing thesis and the universal Turing machine.) Therefore, if any digital machine can "act like it is thinking" then, every sufficiently powerful digital machine can. Turing writes, "all digital computers are in a sense equivalent."^[8]

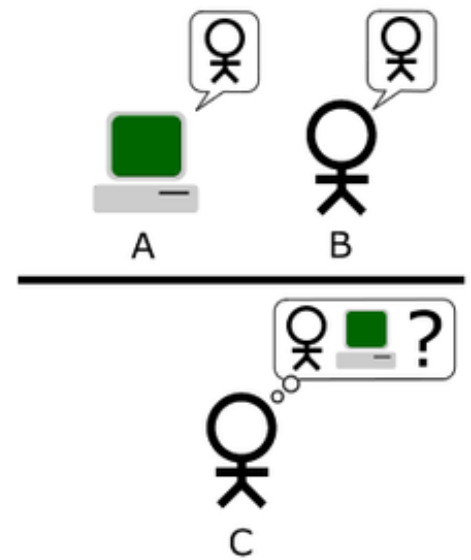
This allows the original question to be made even more specific. Turing now restates the original question as "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"^[8]

Hence Turing states that the focus is not on "whether all digital computers would do well in the game nor whether the computers that are presently available would do well, but whether there are imaginable computers which would do well".^[9] What is more important is to consider the advancements possible in the state of our machines today regardless of whether we have the available resource to create one or not.

Nine common objections

Having clarified the question, Turing turned to answering it: he considered the following nine common objections, which include all the major arguments against artificial intelligence raised in the years since his paper was first published.^[10]

1. **Religious Objection:** This states that thinking is a function of man's immortal soul; therefore, a machine cannot think. "In attempting to construct such machines," wrote Turing, "we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children; rather we are, in either case, instruments of His will providing mansions for the souls that He creates."



The "standard interpretation" of the Turing Test, in which the interrogator is tasked with trying to determine which player is a computer and which is a human

2. *'Heads in the Sand' Objection*: "The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so." This thinking is popular among intellectual people, as they believe superiority derives from higher intelligence and the possibility of being overtaken is a threat (as machines have efficient memory capacities and processing speed, machines exceeding the learning and knowledge capabilities are highly probable). This objection is a fallacious appeal to consequences, confusing what should not be with what can or cannot be (Wardrip-Fruin, 56). → 강요된 생각
3. *Mathematical Objections*: This objection uses mathematical theorems, such as Gödel's incompleteness theorem, to show that there are limits to what questions a computer system based on logic can answer. Turing suggests that humans are too often wrong themselves and pleased at the fallibility of a machine. (This argument would be made again by philosopher John Lucas in 1961 and physicist Roger Penrose in 1989.)^[11] ← 오류를 범하기 쉬운
4. *Argument From Consciousness*: This argument, suggested by Professor Geoffrey Jefferson in his 1949 Lister Oration states that "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."^[12] Turing replies by saying that we have no way of knowing that any individual other than ourselves experiences emotions, and that therefore we should accept the test. He adds, "I do not wish to give the impression that I think there is no mystery about consciousness... [b]ut I do not think these mysteries necessarily need to be solved before we can answer the question [of whether machines can think]." (This argument, that a computer can't have *conscious experiences* or *understanding*, would be made in 1980 by philosopher John Searle in his Chinese room argument. Turing's reply is now known as the "other minds reply". See also Can a machine have a mind? in the philosophy of AI.)^[13]
5. *Arguments from various disabilities*. These arguments all have the form "a computer will never do X". Turing offers a selection:

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, 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.

Turing notes that "no support is usually offered for these statements," and that they depend on naive assumptions about how versatile machines may be in the future, or are "disguised forms of the argument from consciousness." He chooses to answer a few of them:

1. *Machines cannot make mistakes*. He notes it's easy to program a machine to appear to make a mistake.
 2. *A machine cannot be the subject of its own thought* (or can't be self-aware). A program which can report on its internal states and processes, in the simple sense of a debugger program, can certainly be written. Turing asserts "a machine can undoubtably be its own subject matter."
 3. *A machine cannot have much diversity of behaviour*. He notes that, with enough storage capacity, a computer can behave in an astronomical number of different ways.
6. *Lady Lovelace's Objection*: One of the most famous objections states that computers are incapable of originality. This is largely because, according to Ada Lovelace, machines are incapable of independent learning.

The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths. → 허세, 자랑 or 자부심

→ 주장, 권리의 행사

Turing suggests that Lovelace's objection can be reduced to the **assertion** that computers "can never take us by surprise" and argues that, to the contrary, computers could still surprise humans, in particular where the consequences of different facts are not immediately recognizable. Turing also argues that Lady Lovelace was **hampered** by the context from which she wrote, and if exposed to more contemporary scientific knowledge, it would become evident that the brain's storage is quite similar to that of a computer. → 방해하다

7. *Argument from continuity in the nervous system*: Modern neurological research has shown that the brain is not digital. Even though neurons fire in an all-or-nothing pulse, both the exact timing of the pulse and the probability of the pulse occurring have analog components. Turing acknowledges this, but argues that any analog system can be simulated to a reasonable degree of accuracy given enough computing power. (Philosopher Hubert Dreyfus would make this argument against "the biological assumption" in 1972.)^[14]
8. *Argument from the informality of behaviour*: This argument states that any system governed by laws will be predictable and therefore not truly intelligent. Turing replies by stating that this is confusing laws of behaviour with general rules of conduct, and that if on a broad enough scale (such as is **evident** in man) machine behaviour would become increasingly difficult to predict. He argues that, just because we can't immediately see what the laws are, does not mean that no such laws exist. He writes "we certainly know of no circumstances under which we could say, 'we have searched enough. There are no such laws.'" (Hubert Dreyfus would argue in 1972 that human reason and problem solving was not based on formal rules, but instead relied on instincts and awareness that would never be captured in rules. More recent AI research in robotics and computational intelligence attempts to find the complex rules that govern our "informal" and unconscious skills of perception, mobility and pattern matching. See Dreyfus' critique of AI).^[15] This rejoinder also includes the **Turing's Wager** argument.
9. *Extra-sensory perception*: In 1950, extra-sensory perception was an active area of research and Turing chooses to give ESP the benefit of the doubt, arguing that conditions could be created in which mind-reading would not affect the test. → 비기, 실험

Learning machines

In the final section of the paper Turing details his thoughts about the Learning Machine that could play the imitation game successfully.

Here Turing first returns to Lady Lovelace's objection that the machine can only do what we tell it to do and he likens it to a situation where a man "injects" an idea into the machine to which the machine responds and then falls off into **quiescence**. He extends on this thought by an analogy to an atomic pile of less than critical size which is to be considered the machine and an injected idea is to correspond to a neutron entering the pile from outside the pile; the neutron will cause a certain disturbance which eventually **dies away**. Turing then builds on that analogy and mentions that if the size of the pile were to be sufficiently large then a neutron entering the **pile** would cause a disturbance that would continue to increase until the whole pile were destroyed, the pile would be supercritical. Turing then asks the question as to whether this analogy of a supercritical pile could be extended to a human mind and then to a machine. He concludes that such an **analogy** would indeed be suitable for the human mind with "There does seem to be one for the human mind. The majority of them seem to be "subcritical," i.e., to correspond in this analogy to piles of sub critical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion are supercritical. An idea presented to such a mind that may give rise to a whole "theory" consisting of secondary, tertiary and more remote ideas". He finally asks if a machine could be made to be **supercritical**. → 조양제

Turing then mentions that the task of being able to create a machine that could play the imitation game is one of programming and he postulates that by the end of the century it will indeed be technologically possible to program a machine to play the game. He then mentions that in the process of trying to imitate an adult human

mind it becomes important to consider the processes that lead to the adult mind being in its present state; which he summarizes as:

1. The initial state of the mind, say at birth,
2. The education to which it has been subjected,
3. Other experience, not to be described as education, to which it has been subjected.

Given this process he asks whether it would be more appropriate to program a child's mind instead of an adult's mind and then subject the child mind to a period of education. He likens the child to a newly bought notebook and speculates that due to its simplicity it would be more easily programmed. The problem then is broken down into two parts, the programming of a child mind and its education process. He mentions that a child mind would not be expected as desired by the experimenter (programmer) at the first attempt. A learning process that involves a method of reward and punishment must be in place that will select desirable patterns in the mind. This whole process, Turing mentions, to a large extent is similar to that of evolution by natural selection where the similarities are:

Structure of the child machine = hereditary material
Changes of the child machine = mutations
Natural selection = judgment of the experimenter

Following this discussion Turing addresses certain specific aspects of the learning machine:

- Nature of inherent complexity: The child machine could either be one that is as simple as possible, merely maintaining consistency with general principles, or the machine could be one with a complete system of logical inference programmed into it. This more complex system is explained by Turing as "...would be such that the machine's store would be largely occupied with definitions and propositions. The propositions would have various kinds of status, e.g., well-established facts, conjectures, mathematically proved theorems, statements given by an authority, expressions having the logical form of proposition but not belief-value. Certain propositions may be described as "imperatives." The machine should be so constructed that as soon as an imperative is classed as "well established" the appropriate action automatically takes place." Despite this built-in logic system the logical inference programmed in would not be one that is formal, rather it would be one that is more pragmatic. In addition the machine would build on its built-in logic system by a method of "scientific induction".
- Ignorance of the experimenter: An important feature of a learning machine that Turing points out is the ignorance of the teacher of the machine's internal state during the learning process. This is in contrast to a conventional discrete state machine where the objective is to have a clear understanding of the internal state of the machine at every moment during the computation. The machine will be seen to be doing things that we often cannot make sense of or something that we consider to be completely random. Turing mentions that this specific character bestows upon a machine a certain degree of what we consider to be intelligence, in that intelligent behaviour consists of a deviation from the complete determinism of conventional computation but only so long as the deviation does not give rise to pointless loops or random behaviour.
- The importance of random behaviour: Though Turing cautions us of random behaviour he mentions that inculcating an element of randomness in a learning machine would be of value in a system. He mentions that this could be of value where there might be multiple correct answers or ones where it might be such that a systematic approach would investigate several unsatisfactory solutions to a problem before finding the optimal solution which would entail the systematic process inefficient. Turing also mentions that the process of evolution takes the path of random mutations in order to find solutions that

benefit an organism but he also admits that in the case of evolution the systematic method of finding a solution would not be possible.

Turing concludes by speculating about a time when machines will compete with humans on numerous intellectual tasks and suggests tasks that could be used to make that start. Turing then suggests that abstract tasks such as playing chess could be a good place to start another method which he puts as "...it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English."

An examination of the development in artificial intelligence that has followed reveals that the learning machine did take the abstract path suggested by Turing as in the case of Deep Blue, a chess playing computer developed by IBM and one which defeated the world champion Garry Kasparov (though, this too is controversial) and the numerous computer chess games which can outplay most amateurs.^[16] As for the second suggestion Turing makes, it has been likened by some authors as a call to finding a simulacrum of human cognitive development.^[16] And such attempts at finding the underlying algorithms by which children learn of the features of the world around them are only beginning to be made.^{[16][17][18]} 발견, 모형

Notes

1. Turing 1950, p. 433
2. Turing 1950, p. 434
3. This describes the simplest version of the test. For a more detailed discussion, see Versions of the Turing test.
4. Harnad, Stevan (2008), "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence" (<http://eprints.ecs.soton.ac.uk/12954/>), in Epstein, Robert; Peters, Grace (eds.), *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Kluwer
5. Harnad, Stevan (2001), "Minds, Machines, and Turing: The Indistinguishability of Indistinguishables" (<http://cogprints.org/2615/>), *Journal of Logic, Language and Information*, 9 (4): 425–445, doi:10.1023/A:1008315308862 (<https://doi.org/10.1023%2FA%3A1008315308862>), S2CID 1911720 (<https://api.semanticscholar.org/CorpusID:1911720>).
6. Wardrip-Fruin, Noah and Nick Montfort, ed (2003). *The New Media Reader*. The MIT Press. ISBN 0-262-23227-8.
7. Harnad, Stevan (1992), "The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion" (<http://cogprints.org/1584/>), *SIGART Bulletin*, 3 (4): 9–10, doi:10.1145/141420.141422 (<https://doi.org/10.1145%2F141420.141422>), S2CID 36356326 (<https://api.semanticscholar.org/CorpusID:36356326>).
8. Turing 1950, p. 442
9. Turing 1950, p. 436
10. Turing 1950 and see Russell & Norvig 2003, p. 948 where comment "Turing examined a wide variety of possible objections to the possibility of intelligent machines, including virtually all of those that have been raised in the half century since his paper appeared."
11. Lucas 1961, Penrose 1989, Hofstadter 1979, pp. 471–473, 476–477 and Russell & Norvig 2003, pp. 949–950. Russell and Norvig identify Lucas and Penrose's arguments as being the same one answered by Turing.
12. "The Mind of Mechanical Man" (<http://www.turingarchive.org/browse.php/B/44>)
13. Searle 1980 and Russell & Norvig 2003, pp. 958–960, who identify Searle's argument with the one Turing answers.
14. Dreyfus 1979, p. 156

15. Dreyfus 1972, Dreyfus & Dreyfus 1986, Moravec 1988 and Russell & Norvig 2003, pp. 51–52, who identify Dreyfus' argument with the one Turing answers.
16. Epstein, Robert; Roberts, Gary; Beber, Grace (2008). *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (https://books.google.com/books?id=aggUJL_5_oQC). Springer. p. 65. ISBN 978-1-4020-6710-5.
17. Gopnik, Alison; Meltzoff., Andrew N. (1997). *Words, thoughts, and theories* (<http://mitpress.mit.edu/books/words-thoughts-and-theories>). MIT Press.
18. Meltzoff, Andrew N. (1999). "Origins of theory of mind, cognition and communication" (http://ilabs.washington.edu/meltzoff/pdf/99Meltzoff_JCommDisord.pdf) (PDF). *Journal of Communication Disorders*. **32** (4): 251–269. doi:10.1016/S0021-9924(99)00009-X (<https://doi.org/10.1016%2FS0021-9924%2899%2900009-X>). PMC 3629913 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3629913>). PMID 10466097 (<https://pubmed.ncbi.nlm.nih.gov/10466097>).

References

- Brooks, Rodney (1990), "Elephants Don't Play Chess" (<http://people.csail.mit.edu/brooks/paper/s/elephants.pdf>) (PDF), *Robotics and Autonomous Systems*, **6** (1–2): 3–15, CiteSeerX 10.1.1.588.7539 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.7539>), doi:10.1016/S0921-8890(05)80025-9 (<https://doi.org/10.1016%2FS0921-8890%2805%2980025-9>), retrieved 30 August 2007
- Crevier, Daniel (1993), *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, ISBN 0-465-02997-3
- Dreyfus, Hubert (1972), *What Computers Can't Do*, New York: MIT Press, ISBN 978-0-06-011082-6
- Dreyfus, Hubert; Dreyfus, Stuart (1986), *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Oxford, UK: Blackwell
- Dreyfus, Hubert (1979), *What Computers Still Can't Do*, New York: MIT Press.
- Harnad, Stevan; Scherzer, Peter (2008), "First, Scale Up to the Robotic Turing Test, Then Worry About Feeling" (<http://eprints.ecs.soton.ac.uk/14430/>), *Artificial Intelligence in Medicine*, **44** (2): 83–9, CiteSeerX 10.1.1.115.4269 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.4269>), doi:10.1016/j.artmed.2008.08.008 (<https://doi.org/10.1016%2Fj.artmed.2008.08.008>), PMID 18930641 (<https://pubmed.ncbi.nlm.nih.gov/18930641>).
- Haugeland, John (1985), *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: MIT Press.
- Moravec, Hans (1976), *The Role of Raw Power in Intelligence* (<https://web.archive.org/web/20160303232511/http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html>), archived from the original (<http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html>) on 3 March 2016, retrieved 7 November 2007
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (<http://aima.cs.berkeley.edu/>) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2
- Searle, John (1980), "Minds, Brains and Programs" (<http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>), *Behavioral and Brain Sciences*, **3** (3): 417–457, doi:10.1017/S0140525X00005756 (<https://doi.org/10.1017%2FS0140525X00005756>)
- Turing, Alan (October 1950), "Computing Machinery and Intelligence" (<https://www.csee.umbc.edu/courses/471/papers/turing.pdf>) (PDF), *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433 (<https://doi.org/10.1093%2Fmind%2FLIX.236.433>)
- Saygin, A. P. (2000). "Turing Test: 50 years later". *Minds and Machines*. **10** (4): 463–518. doi:10.1023/A:1011288000451 (<https://doi.org/10.1023%2FA%3A1011288000451>). hdl:11693/24987 (<https://hdl.handle.net/11693%2F24987>). S2CID 990084 (<https://api.semanticscholar.org/CorpusID:990084>).

- Noah Wardrip-Fruin and Nick Montfort, eds. (2003). *The New Media Reader*. Cambridge: MIT Press. ISBN 0-262-23227-8. "Lucasfilm's Habitat" pp. 663–677.

External links

- PDF with the full text of the paper (<https://academic.oup.com/mind/article-pdf/LIX/236/433/9866119/433.pdf>)
 - Saygin, Ayse Pinar; Cicekli, Ilyas; Akman, Varol (1999). "An analysis and review of the next 50 years". *Minds and Machines*: 2000. CiteSeerX 10.1.1.157.1592 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.157.1592>).
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Computing_Machinery_and_Intelligence&oldid=991143294"

This page was last edited on 28 November 2020, at 14:13 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.