MIT Press • Open Encyclopedia of Cognitive Science

The Turing Test

Diane Proudfoot[1]

[1]University of Canterbury

MIT Press

Published on: Jul 24, 2024

DOI: https://doi.org/10.21428/e2759450.7678921b

"The Turing test" is the name given to a test of human-level intelligence in machines, invented by Alan Turing, the renowned mathematician, codebreaker, and computer pioneer. In Turing's imitation game, a human interrogator has text conversations with both a human being and a computer that is pretending to be human; the interrogator's goal is to identify the computer. Computers that mislead interrogators often enough, Turing proposes, can *think*. Numerous researchers—most recently, the designers of large language models (LLMs)—have claimed that their models pass Turing's test. These claims reflect the importance of testing to computer science and cognitive science: Without a test, we cannot assess progress toward building human-level AI. As to whether Turing's imitation game is in fact the test that we need, the jury is still out.

History

Turing presented three versions of his imitation game. His 1950 game is the one generally known today: It consists of two simultaneous conversations. Player C (a human) converses with two hidden players, A (a digital computer) and B (a human), and must judge which is which. (Turing's test is not limited to the computer technologies of the 1950s: For example, A could be a large neural network, a DNA computer, or a quantum computer.) The game is unrestricted, in that C can communicate with A and B on "almost any" subject (Turing, 1950, p. 435). Turing said that the game gives us a "criterion" for thinking: If A does well in this game, it can think (Turing, 1950, p. 436). In 1952, he described another version of the game: Each member of a (human) jury interrogates a single hidden player, who may be A or B. Turing pointed out a risk for this version: Interrogators, to avoid the embarrassment of being fooled by a machine, might declare the hidden player a machine every time "without proper consideration" (Turing et al., 1952; Copeland, 2004, p. 495). In 1948, he also described a "little experiment" that was, he said, "a rather idealized form of an experiment I have actually done." In this restricted imitation game, C plays chess with both A, a chess program, and B, a human chess-player. Turing said, "C may find it quite difficult to tell which he is playing" (Turing, 1948; Copeland, 2004, p. 431).

Almost immediately, optimistic researchers declared that their machines passed Turing's test; more recently, well-known bots and AI systems such as Watson, Eugene Goostman, Google Duplex, AlphaStar, LaMDA, and ChatGPT-4 have been said to pass the test [see Large Language Models]. Some of these researchers claimed only that their machines passed a very restricted form of Turing's test, but some declared that their machines could really think; and some researchers said that, just because their (very limited) machines passed Turing's test, the test was broken. All these claims are moot, however, since to date no reported test has conformed to Turing's specified parameters. Turing himself was confident that some machine would pass his unrestricted test, although he predicted in 1952 that this would take at least 100 years.

Core concepts

The unrestricted test that Turing specified in 1950 has five key features. (1) It assesses the machine's ability at *open domain conversation*, in contrast to tests of expert systems, which are confined to a narrow field (e.g., medical diagnosis from laboratory results). (2) If a machine does well in Turing's imitation game, it is deemed to think in the *everyday* sense of "think." (3) If a machine fails to do well in the game, it does not follow that it does *not* think (Turing recognized that an intelligent machine might do badly in the game). For Turing, doing well in the game is a sufficient, but not a necessary, condition for thinking—and therefore not a definition. In fact, he said: "I don't want to give a definition of thinking" (Turing et al., 1952; Copeland, 2004, p. 494). (4) Turing's test is *qualitative* and *discursive*, and it *disallows the tricky questions* that computer scientists have typically used to unmask chatbots—for example, "John Wilkes Booth shot Abraham Lincoln because he was defenseless. Who was defenseless?". In contrast, other AI benchmarks (e.g., the General Language Understanding Evaluation test, which calculates an LLM's capacity at natural-language processing) are quantitative, evaluate each system on a standard set of tasks, and typically compare one machine with another (rather than with a human). (5) Turing's test is *architecture independent*. Turing himself proposed different machine models, including the universal Turing machine and unorganized neural networks, and also suggested building robotic "child machines" as a route to AI (Copeland, 2023).

Questions, controversies, and new developments

Turing did not stipulate a duration for the game; he did not even say that a duration must be set, rather than leaving interrogators to play until confident of their decision. Nor did Turing specify how many times to play the game before judging the machine. There is also considerable debate as to how to score the game. Several accounts take Turing's 1950 prediction concerning progress in AI—namely, that in about 50 years, an average interrogator would have no more than 70% chance of correctly identifying a computer (with a specified storage capacity) after "five minutes of questioning"—as stating the parameters for scoring the game. Other accounts claim that the machine passes if it does better than chance at convincing the interrogator that it is the human. Yet other approaches compare an interrogator's accuracy in the computer-imitates-human game with an interrogator's accuracy in another imitation game that Turing used to introduce his test: Here, C's task is to distinguish a man pretending to be a woman from an actual woman. Turing asked if the interrogator in the computer-imitates-human game would decide wrongly "as often" as the interrogator in this other game (Turing, 1950, p. 434).

Two questions about Turing's test are central. First, *why* did Turing switch from asking "Can machines think?" to asking "Are there imaginable digital computers which would do well in the imitation game?" (Turing, 1950, p. 442). Three different answers are given. On the behaviorist view, there is nothing more to thinking than *behaving as if* you think (Searle, 1980). On the inductive view, success in the game is evidence that the machine can imitate—or on some accounts duplicate—*the relevant features of the brain* of a thinking human (Moor, 1976). On the response-dependence view, a machine thinks if in the game *we*

react to it as such (Proudfoot, 2013). Turing said that intelligence is an "emotional concept" and that whether a machine thinks depends as much on us as on the machine (Turing, 1948; Copeland, 2004, p. 431).

Second, is Turing's test a *satisfactory* test for thinking in machines? Objections to the test (ignoring those that target mistaken or confused accounts of the test) are broadly of four kinds. Turing's test is, it is claimed: (1) *too narrow*—a test for AI should not focus on species-specific intelligence (Ford & Hayes, 1998); (2) *all-or-nothing*—a test for AI should facilitate science's incremental progress toward the goal of a thinking machine; (3) a test only of *outward manifestations* of thinking, not thinking itself, with the result that an imaginary vast lookup-table mechanism could pass (Block, 1981); and (4) *outdated*—very soon, generative AI models will pass Turing's test, despite not thinking. Some researchers even claim that Turing introduced his imitation game, not as a serious test, but solely as a rhetorical device to counter skepticism about AI. Yet, at least some of these objections can be countered. Objection 4 is only the latest in a long line of claims, all so far premature, that a new system renders Turing's test obsolete. Regarding 3, some theorists have argued that it is mathematically impossible for a vast lookup-table mechanism to pass Turing's test in the actual world—and Turing focused on real-world machines. Regarding 2, strictly the test is not all-or-nothing, as the machine's task can be made more or less difficult by varying the characteristics of the human player (e.g., age or linguistic ability). It is unknown, though, if continually adapting the imitation game in this way would help scientists progress from simple chatbots to human-level AI.

Whether objections 1 through 4 in fact undermine Turing's test depends on larger issues. For example, objection 1 raises the following question: By the term "thinking," do we mean a characteristic of both insects and human beings—or are humans the paradigm case of "thinking," just as the original meter stick in Paris was the standard for the meaning of "meter"? And objection 3 raises the following question: By "cognition," do we mean only what occurs inside the "black box" of the brain—or do we also include speaking, writing, or even using a smartphone? How someone answers these more fundamental questions will influence their view of Turing's test.

Broader connections

Various Turing-style tests (replicating only part of Turing's format) have been proposed. Some require a "thinking" machine to imitate, not only human conversation but also human perceptual and motor skills (e.g., Harnad, 1991), physical appearance (e.g., Strathearn & Ma, 2020), or performance on psychometric tests (e.g., Bringsjord, 2011). Some are put forward as tests for more specialized abilities (e.g., social cognition, creativity, empathy, or musical intelligence) or for even grander attributes (e.g., consciousness, personhood, or moral agency). Turing-style tests, including crowd-sourced tests, can also be used to determine if, for example, virtual characters and computer graphics are believable, a social media account is fake, computer diagnosis of radiographic images is reliable, or if an online user is human (in a CAPTCHA, where the judge is a machine).

Some envisaged Turing-style tests involve computer-imitates-animal games to test, for example, if LLM-generated bird calls could fool actual birds. Some have only human players: For example, if a human can put on a convincing display of accepting an ideology (that they do not actually endorse), they are said to understand the ideology, and if a judge cannot distinguish between fMR images of activation (to specific stimuli) in the brains of a healthy subject and of a person in a persistent vegetative state, the person in a vegetative state is held to be conscious.

Turing's test is also linked to ethical questions about AI. For example, it is suggested that the test can be used to decide if a machine is "transparent"—if a machine's explanation (e.g., of some choice it makes) is indistinguishable from a human's explanation, the machine is said to be understandable. Transparent machines, it is hoped, will allay the widespread fear that we cannot trust AI. In addition, Turing's test raises the question: If a machine passes, does it have rights, and do humans have moral obligations to the machine?

Further reading

- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
- Copeland, B. J. (Ed.) (2004). *The essential Turing: Seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Clarendon Press. https://doi.org/10.1093/oso/9780198250791.001.0001
- Copeland, B. J., Bowen, J. P., Sprevak, M., Wilson, R. J., et al. (2017). *The Turing guide*. Oxford University Press. https://doi.org/10.1093/oso/9780198747826.001.0001
- Hodges, A. (2019). Alan Turing. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 Edition). https://plato.stanford.edu/archives/win2019/entries/turing/

References

- Block, N. (1981). Psychologism and Behaviorism. *The Philosophical Review*, *90*(1), 5. https://doi.org/10.2307/2184371 ↩
- Bringsjord, S. (2011). Psychometric artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, *23*(3), 271–277. https://doi.org/10.1080/0952813X.2010.502314 ↩
- Copeland, B. J. (2023). Early AI in Britain: Turing et al. *IEEE Annals of the History of Computing*, *45*(3), 19–31. https://doi.org/10.1109/mahc.2023.3300660 ↩
- Copeland, B. J. (Ed.) (2004). *The essential Turing: Seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Clarendon Press. https://doi.org/10.1093/oso/9780198250791.001.0001 ↩

- Ford, K. M., & Hayes, P. J. (1998). On conceptual wings: Rethinking the goals of artificial intelligence. *Scientific American Presents*, *9*(4), 78–83.

  ↩
- Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, *1*, 43–54. https://doi.org/10.1007/BF00360578

  ↩
- Moor, J. H. (1976). An analysis of the Turing test. *Philosophical Studies*, *30*(4), 249–257. https://doi.org/10.1007/BF00372497

  ↩
- Proudfoot, D. (2013). Rethinking Turing's test. *The Journal of Philosophy*, *110*(7), 391–411. https://doi.org/10.5840/jphil2013110722

  ↩
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457. https://doi.org/10.1017/S0140525X00005756

  ↩
- Strathearn, C., & Ma, E. (2020). The multimodal Turing test for realistic humanoid robots with embodied artificial intelligence. In A. Stein, S. Tomforde, J. Botev & P. Lewis (Eds.), *Joint Proceedings of LIFELIKE 2020 and LIFELIKE 2021*. https://ceur-ws.org/Vol-3007/2020-paper-2.pdf

  ↩
- Turing, A. M. (1948). *Intelligent machinery*. National Physical Laboratory. https://turingarchive.kings.cam.ac.uk/unpublished-manuscripts-and-drafts-amtc/amt-c-11.

  ↩
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

  ↩
- Turing, A. M., Braithwaite, R., Jefferson, G., & Newman, M. (1952, January 14). *Can automatic calculating machines be said to think?* [BBC Radio Broadcast]. Turing Digital Archive. https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-6.

  ↩