

In his famous paper, *Computing Machinery and Intelligence*, Alan Turing proposes a test for assessing the intelligence of a computer system. Essentially, he claims that a machine can be considered to be intelligent if it can successfully imitate a human in a typewritten conversation. This seems reasonable: if a computer and a human cannot be distinguished in a conversation, then it would appear that they must have comparable levels of intelligence. The test also allows a wide variety of subjects to be tested. Turing's example transcript shows that a conversation can test knowledge of subjects as diverse as poetry, arithmetic, and chess; he claims that "the question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour." However, the Turing test is hardly a perfect test for evaluating the intelligence of a system. Because it defines intelligences as the ability to carry out a human-like conversation, it necessarily excludes certain other types of intelligence.

If intelligence is defined by the ability to pass the Turing test, any intelligent system must therefore have a strong natural-language processing ability to parse input and generate output responses. Since the human conversing with the machine could bring up any topic, a system must have a strong foundation of general knowledge in order to pass the test. It cannot simply be knowledgeable in one area. This is quite a significant problem, because most practical systems that apply artificial intelligence have their intelligence restricted to one domain. Consider, for example, an expert system for a medical application. It may have a massively complex set of rules and knowledge, and thus be as effective in its field as a human, or even more effective. One could reasonably claim that it has some intelligence. However, since the system's expertise is limited to one area, it will not be able to carry on a conversation; the Turing test can therefore make no assessment about what intelligence the system may possess. Similarly, much effort has been invested into designing systems to effectively play chess. Modern chess-playing systems have reached a level where they are competitive with highly-ranked human players. It seems reasonable that they can be considered intelligent — perhaps even close to human levels of intelligence. But such systems presumably have neither a natural-language interface nor knowledge of other fields, so they will have limited success under Turing's criteria. The Turing test is thus ineffective for evaluating the vast majority of systems, which are only partially intelligent, specializing in one particular domain.

The Turing test is also necessarily biased towards systems that are particularly capable of performing the task it tests for: feigning human responses. Some of the most effective programs in this regard have been relatively simple rule-based systems such as Weizenbaum's ELIZA program and other similar programs. Another strategy is to employ a database of responses to the most common queries. These programs are neither especially practical nor of human-level intelligence, but they can perform reasonably well at the Turing test. Even the most advanced systems of this type would have difficulty convincing an interrogator that they are human. However, this reveals another flaw in the Turing test's methodology: the results do not just depend on the computer being, but also the two humans acting as interrogator and the other test subject. Whether a system

is considered intelligent or not can depend on what questions are posed to it: whether it receives questions it is prepared to answer, or questions that are particularly effective at revealing it to be a machine. Furthermore, the Turing test also requires that any intelligent system be human-like in its characteristics. Turing notes that a machine could be “unmasked because of its deadly accuracy” and thus would have to “deliberately introduce mistakes” in order to seem human. This leads to the conclusion that a machine that gives wrong answers occasionally is more intelligent because it is more human. The Turing test does not account for intelligence that is not equivalent to human intelligence. It can be an effective experiment to identify a certain type of intelligence, but fails in the broader sense.