

The Winograd Schema Challenge *by Hector J. Levesque, Ernest Davis, and Leora Morgenstern***Summary:**

The Turing Test named Imitation game proposed by Alan Turing focuses on observable behavior where the machine must produce a behavior that is required thought in people. In this test, if an interrogator were unable to tell after a long, free-flowing and unrestricted conversation with a machine whether she was talking to a person or a machine, then we would say that machine is thinking. Author, completely agree that English text is a sufficient medium for displaying intelligent behavior. But, author slightly disagrees that free-flowing conversation as a right sort of test. The free-flowing conversation has a chance of deception and trickery. So, in this paper author proposed the Winograd Schema (WS) Challenge which is an alternative test to Turing Test that has some conceptual and practical advantages. In WS challenge machine will not be imitating humans, where it is a pair of sentences differing in only one or two words containing an ambiguity that is resolved in opposite ways in the two sentences and that requires the use of world knowledge and reasoning for its resolution. The machine would have to resolve the ambiguity by giving binary input. To pass the challenge, the machine would have to achieve near-human levels of success; presumably close to 100%. This paper is concerned about following questions:

- How is machine supposed to deal with this evasiveness and determine if there is any comprehension behind the verbal acrobatics?
- How can we steer research in a more constructive direction, away from deception and trickery?

To deal with evasiveness few researchers have taken more time in processing the given text but in challenges, the grading must be completed in 5 minutes which is problematic. To identify the deception and trickery there was an idea of captcha which is a distorted image of multidigit number where the human can easily identify it. But these methods were just able to identify if a machine is talking to the human or a machine. The behavior of the approaches may be ideal but the scope of what are asking to do may be limited. Recognizing Textual Entailment (RTE) which is a new type of Turing Test where a subject is presented with a series of yes-no questions concerning whether one English sentence entails other gave an approach for authors to construct WS challenge. The WS proposed in this paper is a variant of the RTE that requires subjects to answer binary questions and appeal to world knowledge and default reasoning abilities, but without depending on an explicit notion of entailment. In each question in WS challenge, it must follow these features: (1) It should be easily disambiguated by a human reader. Ideally, this should be so easy that the reader does not even notice that there is an ambiguity. (2) It should not be solvable by simple techniques such as selectional restrictions. (3) It should be Google-proof; that is, there should be no obvious statistical test over text corpora that will reliably disambiguate these correctly. The main challenges on the way to this challenge are (1) we want normally-abled adults whose first language is English to the answers obvious. But what do we mean by obvious? (2) these are questions where the choice between the two parties can be made without considering the relationship between them expressed by the sentence. (3) sentences can possibly have multiple answers which can lead to statistical differences. To overcome these challenges author chooses examples with randomly chosen proper names of people where there is no change of the names to a special word or its alternative. And the author is also including extra information that makes apparent to the subject. A major advantage of the WS challenge is that it allows incremental progress. In addition, WS can be grouped according to the domain. Even though the proposed challenged in this paper is certainly less demanding than an intelligent conversation imagined by Turing, but it does offer a test challenge that is less subject to abuse.