My Computer Is an Honor Student — but How Intelligent Is It? Standardized Tests as a Measure of AI

Peter Clark
Allen Institute for Artificial Intelligence

Oren Etzioni
Allen Institute for Artificial Intelligence

DOI: https://doi.org/10.1609/aimag.v37i1.2636

Abstract

Given the well-known limitations of the Turing Test, there is a need for objective tests to both focus attention on, and measure progress towards, the goals of AI. In this paper we argue that machine performance on standardized tests should be a key component of any new measure of AI, because attaining a high level of performance requires solving significant AI problems involving language understanding and world modeling - critical skills for any machine that lays claim to intelligence. In addition, standardized tests have all the basic requirements of a practical test: they are accessible, easily comprehensible, clearly measurable, and offer a graduated progression from simple tasks to those requiring deep understanding of the world. Here we propose this task as a challenge problem for the community, summarize our state-of-the-art results on math and science tests, and provide supporting datasets

Standardized Tests as a Measure of AI Peter Clark and Oren Etzioni Abstract Given the well-known limitations of the Turing Test, there is a need for objective tests to both focus attention on, and measure progress towards, the goals of AI. This continuum is highly desirable, as it means that there is a low barrier to entry, allowing researchers to make initial inroads into the task, while significant AI challenges need to be solved to do well in the exam. The diversity of questions also ensures a variety of skills are tested for, and guards against finding a simple shortcut that may answer them all without requiring any depth of understanding. The student feels the objects but does not look into the bag. Which property of the objects can the student most likely identify? Standardized Tests as a Measure of AI. By Clark, Peter; Etzioni, Oren. Read preview. Magazine article AI Magazine. In this article, we argue that standardized tests are an effective and practical assessment of many aspects of machine intelligence, and should be part of any

comprehensive measure of AI progress. While a crisp definition of machine intelligence remains elusive, we can enumerate some general properties we might expect of an intelligent machine. The list is potentially long (for example, Legg and Hutter [2007]), but should at least include the ability to (1) answer a wide variety of questions, (2) answer complex questions, (3) demonstrate commonsense and world knowledge, and (4) acquire new kno Intelligent and intelligence are not quite the same thing. One relates to application, the other to potential. That is why possessing a high IQ, although it would imply that they are intelligent, really implies potential. Unless they apply that IQ then it just remains a potential. However, these tests do not predict how an individual will utilize the potential with which they were born. Some may maximize it, others may squander it. For instance, we can measure how big is a person's short-term memory capacity is by asking them to read strings of digits or words and then repeat them from memory. Still the results of that test could vary based on the physical condition of an individual at the moment of testing (cold, hot, hungry, etc.). Most standardized tests are created to evaluate the information individual students have retained in core competency areas of learning. Although many students tend to find classes like art, music, and physical education to be the most enjoyable at school, there is also a need for reading, writing, and arithmetic. 2. It can be very expensive for some states. Standardized testing is an assessment tool that is reasonably affordable for most geographic locations. New York State, for example, spends just $7 per student on standardized tests. Oregon spends $13 per student, while Georgia spends $14 per student. Measuring them through standardized testing allows teachers, parents, and administrators to understand where each child is at on their own learning path.