

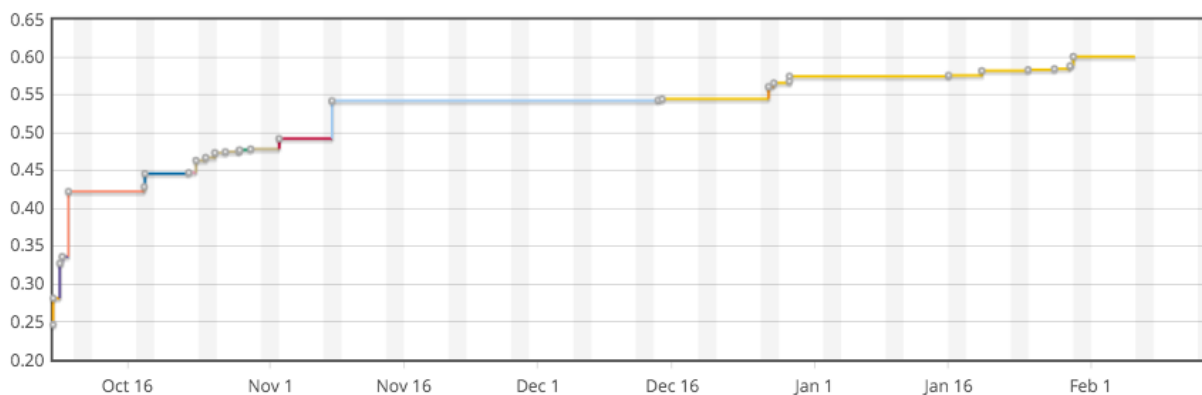
“We’re Not Playing Games” -- Allen AI Introduces Standardized Tests to “Grade” Artificial Intelligence

Seattle, WA (Feb. 16, 2016) -- Today, AI2 announced the results of [The Allen AI Science Challenge](#), which invited scientists worldwide to build AI software that could take a standard 8th grade science test. The goal of the challenge was to assess the state of the art in natural language understanding and reasoning by determining how accurately the participants’ models could answer 8th grade science questions.

“The Allen AI Science Challenge is an important step towards a rational, quantitative assessment of AI’s capabilities, and how these progress over time,” said Dr. Oren Etzioni, CEO at AI2. In contrast with recent AI work on the game ‘Go’ and on Computer Poker, The Science Challenges assesses AI systems in natural-language understanding and knowledge-based reasoning, not just whether a computer can beat a human at a given game.

Over 780 teams participated in the challenge, which lasted four months from October 7th, 2015 through February 13th, 2016. The team achieving the highest-scoring results on the test will be receiving an award of \$50K, with \$20K and \$10K awards for the next best teams. The top teams reached scores near 60% on the final test set of questions. The most successful systems used carefully curated information from science texts and other public resources, which was then searched over using carefully tuned information retrieval techniques to locate the best candidate answer for each multiple choice question.

The leaderboard shown below tracks the top scores over the course of the competition:



Measuring AI: Why science exam questions?

The classical Turing test for AI proposes that if a system appears to exhibit intelligent behavior indistinguishable from that of a human during a natural-language conversation, it could be considered truly “artificially intelligent.” This approach is very game-able, however, and in dire need of revisiting. The New York Times’ John Markoff noted that “the Turing test is a test of

human gullibility.” The “[Beyond the Turing Test](#)” workshop held at the AAAI conference in January of 2015 also took steps toward engaging the community to provide input on the eventual replacement tests for better assessing the success of a given AI system.

A few example questions from the contest highlight the interesting nuances of language and types of reasoning an AI system might need to accomplish in order to successfully produce an answer:

Which part of the eye does light hit first?

- (A) the retina
- (B) the lens
- (C) the cornea
- (D) the pupil

Some types of fish live most of their adult lives in salt water but lay their eggs in freshwater. The ability of these fish to survive in these different environments is an example of

- (A) selective breeding
- (B) learning a new habit
- (C) adaptation
- (D) developmental stages

The Allen Institute for Artificial Intelligence: AI for the Common Good

The Allen Institute for Artificial Intelligence ([AI2](#)) is dedicated to the mission of AI for the common good; building and sharing resources and tools with the wider community to help advance the field of AI in several important areas. AI2 is interested in continuing to develop better ways to measure true progress in the field of artificial intelligence. This means designing tests that are more objective, more understandable, and more applicable to the global challenges we face.

About AI2

AI2 was founded in 2014 with the goal of conducting high-impact research and engineering in the field of artificial intelligence, all for the common good. AI2 is the creation of Paul Allen, Microsoft co-founder, and is led by Dr. Oren Etzioni, a leading researcher in the field of AI. AI2 employs more than 45 top-notch researchers and engineers.

Media Contacts:

Sarah Sullivan

The OutCast Agency

415.823.4351

ssullivan@theoutcastagency.com