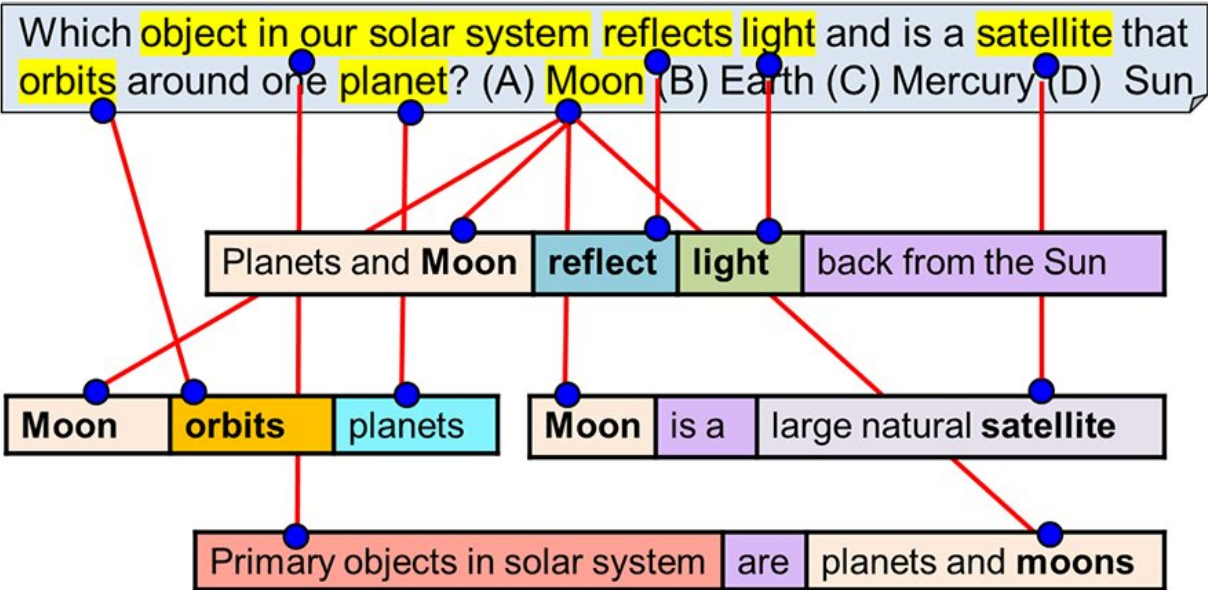


AI Aristo takes science test, emerges multiple-choice superstar

September 6 2019, by Nancy Cohen



Credit: aristo

[Aristo](#) has passed an American eighth grade science test. If you are told Aristo is an earnest kid who loves to read all he can about Faraday and plays the drums you will say so what, big deal.

Aristo, though, is an [artificial intelligence](#) program and scientists would like the world to know this is a big deal, as "a benchmark in AI development," as Melissa Locker called it in *Fast Company*.

We mean, just think about it. Cade Metz, in *The New York Times*, has thought about it. "Four years ago, more than 700 computer [scientists](#) competed in a contest to build artificial intelligence that could pass an eighth-grade science test. There was \$80,000 in prize money on the line. They all flunked. Even the most sophisticated system couldn't do better than 60% on the test. AI couldn't match the language and logic skills that students are expected to have when they enter [high school](#)."

So who is behind the test that in 2019 finally impressed? Not a bad guess: The Allen Institute for Artificial Intelligence, which is overseen by Oren Etzioni. Their system had the [correct answers](#) for more than 90 percent of questions on the test, and it doesn't stop there—the system got over 80 percent of the correct answers on non-diagram multiple choice questions in a 12th grade science exam.

We're now looking at "significant progress in developing AI that can understand languages and mimic the logic and decision-making of humans," said Metz.

For the direct story, you should read "From 'F' to 'A' on the N.Y. Regents Science Exams: An Overview of the Aristo Project," which is now up on [arXiv](#). This project was a six-year mission to answer grade-school and high-school science exams.

The authors were well aware that AI had not made an impressive show in the past of performing on desired levels. With all of AI's mastery at Go, Poker and jeopardy, they said, "the rich variety of standardized exams has remained a landmark challenge. Even in 2016, the best AI system achieved merely 59.3% on an 8th Grade science exam challenge."

The AI took on multiple choice tests; the 90 percent number was on the [exam](#)'s non-diagram, multiple choice questions.

Here is the way the AI2 describe its non-human whiz: "Aristo brings together machine reading and NLP, textual entailment and inference, reasoning with uncertainty, statistical techniques over large corpora, and diagram understanding to develop the first "knowledgeable machine" about science."

The team pampered Aristo for an ulterior motive, less to do with patting themselves on the back and more about what they could learn from Aristo's behaviors on science exams, "as these questions test many of the key skills required for machine intelligence," they said.

In their paper, they explained more about good reasons to leverage standardized science exams.

"Standardized tests, in particular science exams, are a rare example of a challenge that meets these requirements. While not a full [test](#) of machine intelligence, they do explore several capabilities strongly associated with intelligence, including language understanding, reasoning, and use of common-sense knowledge. One of the most interesting and appealing aspects of science exams is their graduated and multifaceted nature; different questions explore different types of knowledge, varying substantially in difficulty. For this reason, they have been used as a compelling—and challenging—task for the field for many years."

New bragging rights: Aristo, the authors said, is the first system to achieve a score of over 90 percent on the non-diagram, multiple choice part of the New York Regents 8th Grade Science Exam.

Stephen Johnson in *Big Think* wrote about Aristo's inability to do diagrams. He said "the system is designed only to [interpret](#) language,

meaning it can answer multiple choice questions, but not those featuring an illustration or graph."

Nonetheless, the performance showed that "modern NLP methods can result in mastery of this task."

For the institute, Aristo's feat is not taken as a perch on the mountain but rather a step in a desired direction. They call it a milestone "on the long road toward a machine that has a deep understanding of science and achieves Paul Allen's original dream of a Digital Aristotle."

More information: allenai.org/aristo/

From 'F' to 'A' on the N.Y. Regents Science Exams: An Overview of the Aristo Project, arXiv:1909.01958 [cs.CL] arxiv.org/abs/1909.01958

© 2019 Science X Network

Citation: AI Aristo takes science test, emerges multiple-choice superstar (2019, September 6) retrieved 13 August 2024 from <https://techxplore.com/news/2019-09-ai-aristo-science-emerges-multiple-choice.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.