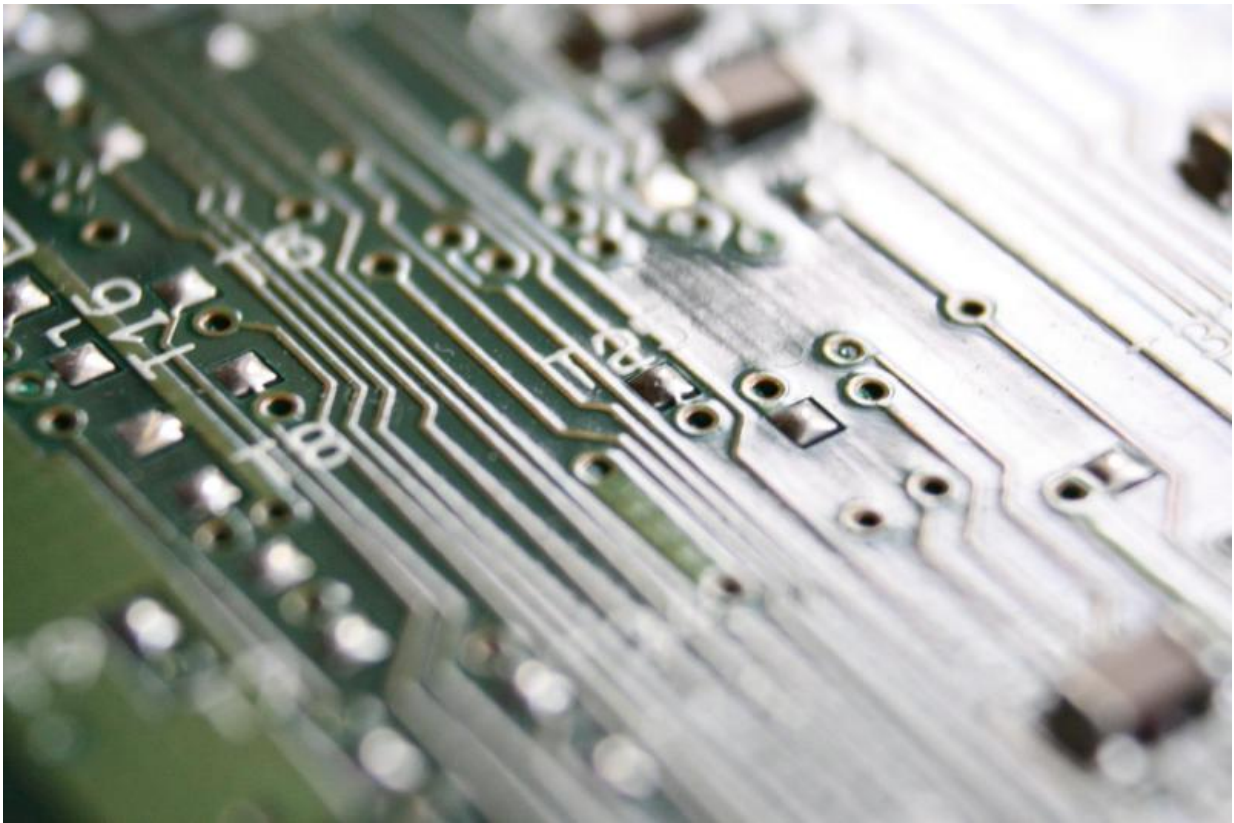# AI machine achieves IQ test score of young child

October 6 2015, by Nancy Owano

Credit: Public Domain

Some people might find it enough reason to worry; others, enough reason to be upbeat about what we can achieve in computer science; all await the next chapters in artificial intelligence to see what more a

machine can do to mimic human intelligence. We already saw what machines can do in arithmetic, chess and pattern recognition.

MIT Technology Review poses the bigger question: to what extent do these [capabilities](#) add up to the equivalent of [human intelligence](#)? Shedding some light on AI and humans, a team went ahead to subject an AI system to a standard IQ test given to humans.

Their paper describing their findings has been posted on arXiv. The team is from the University of Illinois at Chicago and an AI research group in Hungary. The AI system which they used is ConceptNet, an open-source project run by the MIT Common Sense Computing Initiative.

Results: It scored a WPPSI-III VIQ that is average for a four-year-old child, but below average for 5 to 7 year-olds

"We found that the WPPSI-III VIQ psychometric test gives a WPPSI-III VIQ to ConceptNet 4 that is equivalent to that of an average four-year old. The performance of the system fell when compared to older children, and it compared poorly to seven year olds."

They wrote, "In the work reported here, we used the March 2012 joint release of ConceptNet 4 implemented as the Python module conceptnet and AnalogySpace implemented as the Python module divisi2.3. In this paper 'ConceptNet' refers to this combination of AnalogySpace and ConceptNet 4 unless explicitly stated otherwise."

The title of their paper is "Measuring an Artificial Intelligence System's Performance on a Verbal IQ Test For Young Children," and the authors are Stellan Ohlsson, Robert Sloan, György Turán and Aaron Urasky. They represent academic disciplines of statistics, computer science and psychology.

The Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III), which is the test they used, is for children ages 2 years and 6 months to 7 years and 3 months, and is made up of 14 subtests.

The test is called Wechsler after David Wechsler, PhD, cognitive psychology pioneer. Wechsler described intelligence as "the global capacity of a person to act purposefully, to think rationally, and to deal effectively with his environment."

As for the computer's ability to answer questions successfully, the authors discussed the limitations.

An example: Saw was taken as the past tense of see rather than as a cutting tool. "ConceptNet does little or no word-sense disambiguation. It combines different forms of one word into one database entry, to increase what is known about that entry. The lack of disambiguation hurts when, for example, the system's natural language processing tools convert *saw* into the base form of the verb *see*, and our question 'What is a saw used for?' is answered by 'An eye is used to see.'"

The authors said that "In general, more powerful natural language processing tools would likely improve system performance."

Interestingly, these limitations do not spell doom for computers reaching human thought level but rather the limitations help elucidate what needs to come next in AI progress.

MIT Technology Review made the observation that, "Of course, there are various ways that the test could be improved."

Giving the computer natural language processing capabilities is one way. "That would reduce its reliance on the programming necessary to enter the questions and is something that is already becoming possible with

online assistants such as Siri, Cortana, and Google Now," said the report.

MIT Technology Review added this to the bigger picture regarding this IQ study: "Taking Ohlsson and co's result at face value, it's taken 60 years of AI research to build a machine in 2012 that can come anywhere close to matching the common sense reasoning of a four-year old. But the nature of exponential improvements raises the prospect that the next six years might produce similarly dramatic improvements. So a question that we ought to be considering with urgency is: what kind of AI machine might we be grappling with in 2018?"

More information: Measuring an Artificial Intelligence System's Performance on a Verbal IQ Test For Young Children, arXiv:1509.03390 [cs.AI] arxiv.org/abs/1509.03390

Abstract
We administered the Verbal IQ (VIQ) part of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) to the ConceptNet 4 AI system. The test questions (e.g., "Why do we shake hands?") were translated into ConceptNet 4 inputs using a combination of the simple natural language processing tools that come with ConceptNet together with short Python programs that we wrote. The question answering used a version of ConceptNet based on spectral methods. The ConceptNet system scored a WPPSI-III VIQ that is average for a four-year-old child, but below average for 5 to 7 year-olds. Large variations among subtests indicate potential areas of improvement. In particular, results were strongest for the Vocabulary and Similarities subtests, intermediate for the Information subtest, and lowest for the Comprehension and Word Reasoning subtests. Comprehension is the subtest most strongly associated with common sense. The large variations among subtests and ordinary common sense strongly suggest that the WPPSI-III VIQ results do not show that "ConceptNet has the verbal abilities a four-year-old." Rather, children's IQ tests offer one objective metric for the evaluation

and comparison of AI systems. Also, this work continues previous research on Psychometric AI.

© 2015 Tech Xplore