ChatGPT better than undergraduates at solving SAT problems, study suggests

ChatGPT can solve problems at a level that matches or surpasses an undergraduate student, according to a new study. Researchers found that the GPT-3 large language model that underpins the chatbot performed about as well as US college undergraduates when asked to solve reasoning problems that appear on intelligence tests or exams such as the American college admission test, the SAT. Psychologists at the University of California, Los Angeles tested GPT-3's ability to predict the next image in a complex array of shapes, after converting the images to a text format that the model could process and also ensuring the model would never have encountered the questions before. The same problems were put to 40 UCLA undergraduates and the researchers found that GPT-3 solved 80% of the problems correctly, well above the average score of just below 60% for the human participants. The researchers also prompted the model to solve some SAT "analogy" questions – selecting pairs of words that are linked in some way – that they believe had not been published on the internet and therefore could not have appeared in the vast amount of data it was trained on. When compared with college applicants' SAT scores, the UCLA team found that the AI outperformed the average score for humans. In another test the model did less well. The researchers asked it and the student volunteers to match a passage of prose with a different short story that conveyed the same meaning. In this test GPT-3 did less well than the students, although GPT-4 – the improved successor to GPT-3 – did better than its predecessor, according to the research, which was published in the Nature Human Behaviour journal. The study found that GPT-3 displayed a "surprisingly strong" capacity for spotting patterns and inferring relationships, "matching or even surpassing human capabilities in most settings". The study's lead author, Taylor Webb, said that the model driving ChatGPT was not at the standard of artificial general intelligence, or human-level intelligence. He said it struggled with social interactions, mathematical reasoning and solving problems that require understanding physical space, such as working out which tools are best for transferring sweets from one bowl to another. Nonetheless, the technology had made a jump in progress. "It's definitely not fully general human-level intelligence. But it has definitely made progress in a particular area," said Webb, a postdoctoral researcher in psychology at UCLA. The UCLA researchers added that without access to the inner workings of GPT-3, which is developed by the San Francisco-based company OpenAI, they could not determine how the model's reasoning abilities work and whether it is thinking like a human or showing a new form of intelligence. "GPT-3 might be kind of thinking like a human," said Keith Holyoak, a UCLA psychology professor "But on the other hand, people did not learn by ingesting the entire internet, so the training method is completely different. We'd like to know if it's really doing it the way people do, or if it's something brand new – a real artificial intelligence – which would be amazing in its own

right."