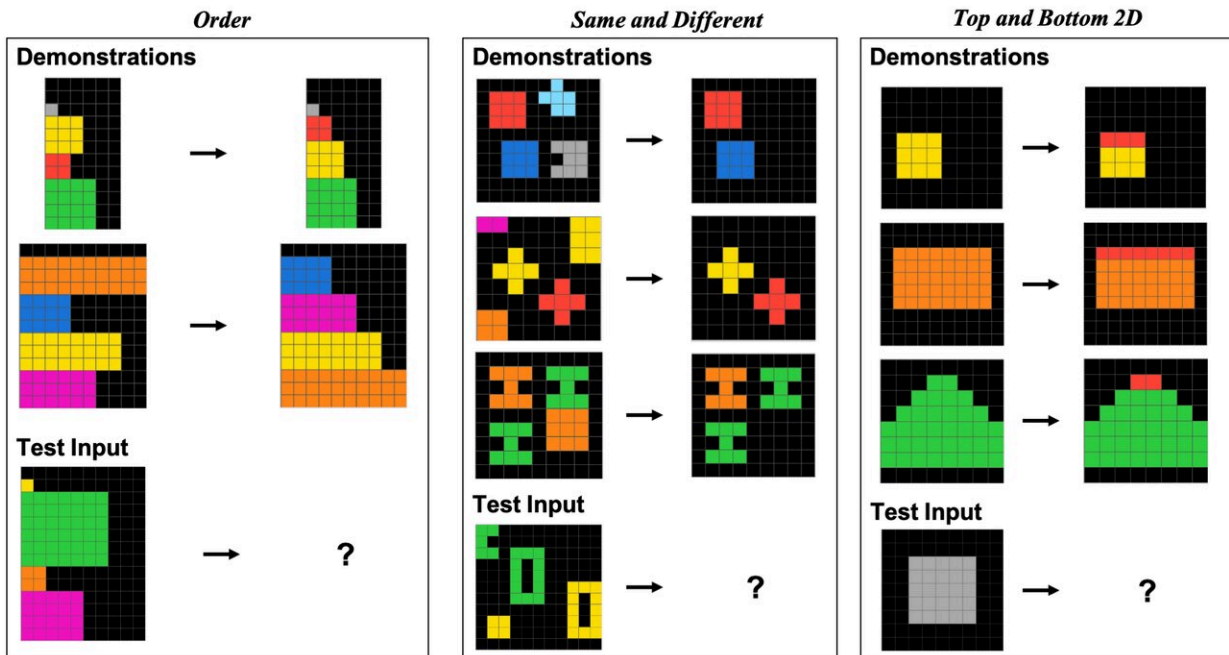


Study: Visual analogies for AI

September 26 2023

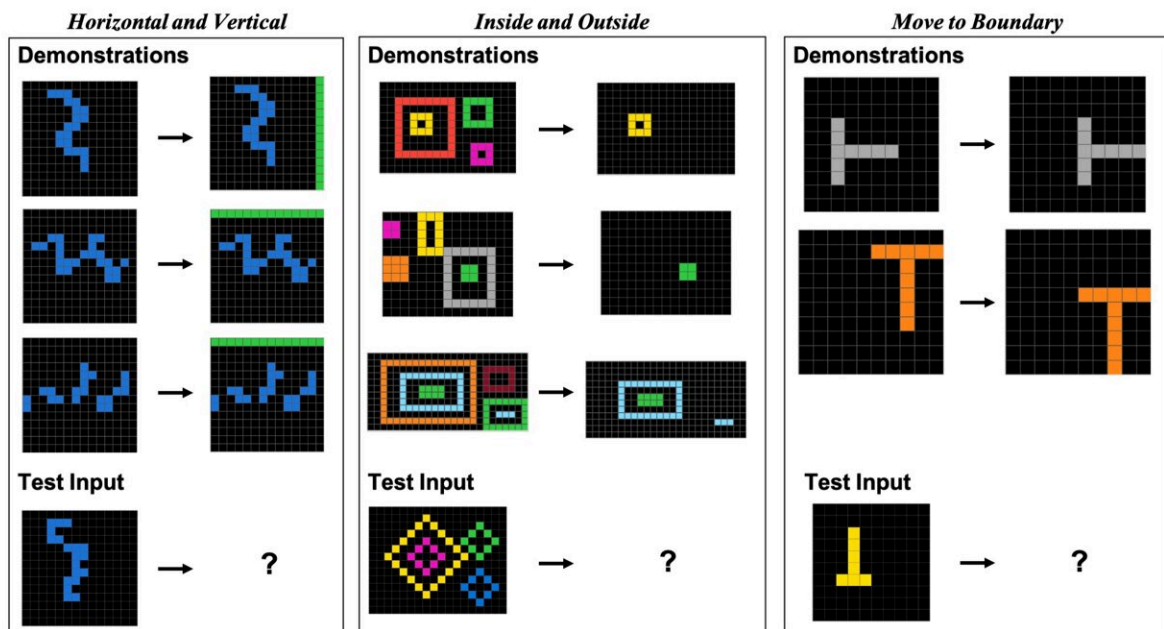


Examples of some of the visual tasks to test intelligence in humans vs AI. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2305.07141

The field of artificial intelligence has long been stymied by the lack of an answer to its most fundamental question: What is intelligence, anyway? AIs such as GPT-4 have highlighted this uncertainty: some researchers believe that GPT models are showing glimmers of genuine intelligence but others disagree.

To address these arguments, we need concrete tasks to pin down and test the notion of [intelligence](#), argue SFI researchers Arseny Moskvichev, Melanie Mitchell, and Victor Vikram Oudouard in [a new paper](#) scheduled for publication in *Transactions on Machine Learning Research*, and posted to the *arXiv* preprint server. The authors provide just that—and find that even the most advanced AIs still lag far behind humans in their ability to abstract and generalize concepts.

The team created evaluation puzzles—based on a domain developed by Google researcher François Chollet—that focus on visual analogy-making, capturing basic concepts such as above, below, center, inside, and outside. Human- and AI test-takers were shown several patterns demonstrating a concept and then asked to apply that concept to a different image. The figure below shows tests of the notion of sameness.



Examples of some of the visual tasks the authors of a recent study posed in testing intelligence. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2305.07141

These visual puzzles were very easy for humans: For example, they got the notion of sameness correct 88% of the time. But GPT-4 struggled, only getting 23% of these puzzles right. So the researchers conclude that currently, AI programs are still weak at visual abstract reasoning.

"We reason a lot by analogies, so that's why it's such an interesting question," Moskvicev says. The team's use of novel visual puzzles ensured that the machines hadn't encountered them before. GPT-4 was trained on large portions of the internet, so it was important to avoid anything it might have encountered already, to be certain it wasn't just parroting existing text rather than demonstrating its own understanding. That's why recent results like an AI's ability to score well on a Bar exam aren't a good test of its true intelligence.

The team believes that as time goes on and AI algorithms improve, developing evaluation routines will get progressively more difficult and more important. Rather than trying to create one test of AI intelligence, we should design more carefully curated datasets focusing on specific facets of intelligence. "The better our algorithms become, the harder it is to figure out what they can and can't do," Moskvicev says. "So we need to be very thoughtful in developing evaluation datasets."

More information: Arseny Moskvicev et al, The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain, *arXiv* (2023). [DOI: 10.48550/arxiv.2305.07141](https://doi.org/10.48550/arxiv.2305.07141)

Provided by Santa Fe Institute

Citation: Study: Visual analogies for AI (2023, September 26) retrieved 27 April 2024 from <https://techxplore.com/news/2023-09-visual-analogies-ai.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.