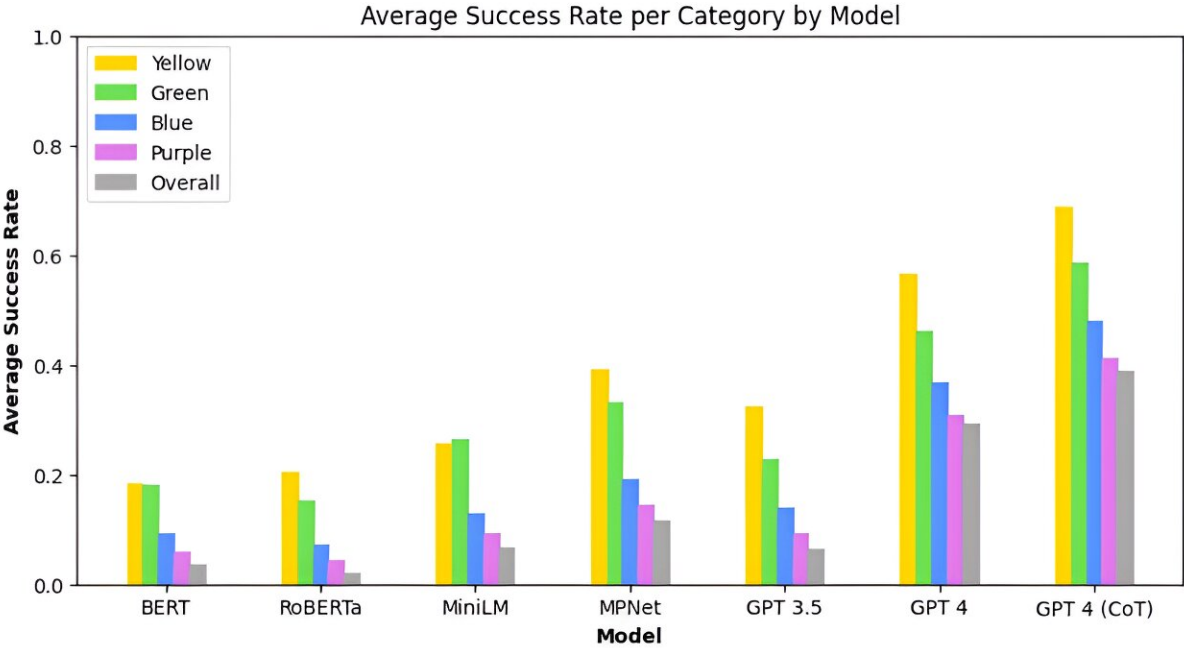


Researchers test AI systems' ability to solve the New York Times' connections puzzle

May 10 2024



Average success rate across all puzzles and seeds for baseline models and LLMs, broken down by puzzle category (note that CoT indicates the use of chain-of-thought prompting). Categories increase in difficulty going from yellow to green to blue to purple. Credit: *arXiv* (2024). DOI: 10.48550/arxiv.2404.11730

Can artificial intelligence (AI) match human skills for finding obscure connections between words? Researchers at NYU Tandon School of Engineering turned to the daily Connections puzzle from The New York

Times to find out.

Connections gives players five attempts to group 16 words into four thematically linked sets of four, progressing from "simple" groups generally connected through straightforward definitions to "tricky" ones reflecting abstract word associations requiring unconventional thinking.

In a study that will be presented at the IEEE 2024 Conference on Games, taking place in Milan, Italy from August 5 to 8, the researchers investigated whether modern natural language processing (NLP) systems could solve these language-based puzzles. The findings are also [published](#) on the *arXiv* preprint server.

With Julian Togelius, NYU Tandon Associate Professor of Computer Science and Engineering (CSE) and Director of the Game Innovation Lab, as the study's senior author, the team explored two AI approaches. The first leveraged GPT-3.5 and recently-released GPT-4, powerful large language models (LLMs) from OpenAI, capable of understanding and generating human-like language.

The second approach used sentence embedding models, namely BERT, RoBERTa, MPNet, and MiniLM, which encode semantic information as vector representations but lack the full language understanding and generation capabilities of LLMs.

The results showed that while all the AI systems could solve some of the Connections puzzles, the task remained challenging overall. GPT-4 solved about 29% of puzzles, significantly better than the embedding methods and GPT-3.5, but far from mastering the [game](#). Notably, the models mirrored [human performance](#) in finding the difficulty levels aligned with the puzzle's categorization from "simple" to "tricky."

"LLMs are becoming increasingly widespread, and investigating where

they fail in the context of the Connections [puzzle](#) can reveal limitations in how they process semantic information," said Graham Todd, Ph.D. student in the Game Innovation Lab who is the study's lead author.

The researchers found that explicitly prompting GPT-4 to reason through the puzzles step-by-step significantly boosted its performance to just over 39% of puzzles solved.

"Our research confirms prior work showing this sort of 'chain-of-thought' prompting can make language models think in more structured ways," said Timothy Merino, Ph.D. student in the Game Innovation Lab who is an author on the study. "Asking the [language](#) models to reason about the tasks that they're accomplishing helps them perform better."

Beyond benchmarking AI capabilities, the researchers are exploring whether models like GPT-4 could assist humans in generating novel word puzzles from scratch. This creative task could push the boundaries of how machine learning systems represent concepts and make contextual inferences.

The researchers conducted their experiments with a dataset of 250 puzzles from an online archive representing daily puzzles from June 12, 2023, to February 16, 2024.

Along with Togelius, Todd and Merino, Sam Earle, a Ph.D. student in the Game Innovation Lab, was also part of the research team. The study contributes to Togelius' body of work that uses AI to improve games and vice versa. Togelius is the author of the 2019 book *Playing Smart: On Games, Intelligence, and Artificial Intelligence*.

More information: Graham Todd et al, Missed Connections: Lateral

Thinking Puzzles for Large Language Models, *arXiv* (2024). [DOI: 10.48550/arxiv.2404.11730](https://doi.org/10.48550/arxiv.2404.11730)

Provided by NYU Tandon School of Engineering

Citation: Researchers test AI systems' ability to solve the New York Times' connections puzzle (2024, May 10) retrieved 23 June 2024 from <https://techxplore.com/news/2024-05-ai-ability-york-puzzle.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.