Sparks of Artificial General Intelligence

A report on early experiments with GPT-4 (Part 1)

About the paper :

This paper is a report of experiments on early versions of OpenAi's GPT-4. It discusses the improving abilities and implications of the recent cohort of LLMs which exhibit more general intelligence than the previous cohort of AI models. It demonstrates that GPT-4 can perform a variety of tasks across a wide range of fields. It discusses the limitations of GPT - 4 and challenges ahead for advancing AGI. It concludes with societal implications of the recent leap in LLMs and future research directions.

About GPT - 4 and LLMs

Large Language Models(LLMs) are neural network models based on the transformer architecture usually trained on a large corpus of web - text data.

LLMs have a self supervised objective of predicting the next word in a partial sentence.

GPT -4 is a Large Language Model recently developed by OpenAI. It was trained using a huge amount of computation and data.

GPT - 4 exhibits many traits of intelligence.

It’s abilities span a wide variety of domains including abstraction, comprehension, vision, coding, mathematics, medicine, law, understanding of human motives and emotions,

Traditional Experimental Approach

- Standard approach in Machine Learning is to evaluate the system on a set of standard benchmark datasets.
- The aim of this approach is to ensure that the model is independent of training data.
- This approach is designed to separate true learning from memorization.

Proposed Experimental Approach

- The traditional approach of evaluating AI models cannot be applied to evaluation of GPT - 4 because of two main reasons :
  - No access to full details of it’s vast training data which means it has potentially seen all existing benchmarks or similar data.
  - One of the key aspects of GPT - 4s intelligence is the ability to perform tasks that go beyond the typical scope of narrow AI systems. GPT - 4 performs the best on tasks which donot have a single solution. In this case evaluation becomes a challenge

Proposed Experimental Approach

- The study proposes an approach closer to traditional psychology.
- The aim of this approach is to generate tasks which demonstrate that GPT - 4s abilities are not just memorization and that it has a deep and flexible understanding of concepts, skills, and domains
- The study probes GPT-4’s responses and behaviors, to verify its consistency, coherence, and correctness and aims to uncover it’s limitations and biases.

Design Of Experiment

In this study, the outlined approach has been executed on the following selected topics to probe the abilities of GPT - 4 :

- Multimodal and interdisciplinary composition
- Coding
- Mathematical Abilities
- Interaction with the world
- Interaction with humans
- Discriminative capabilities

Multimodal and Interdisciplinary composition

Integrative Ability

- To evaluate the model's integrative ability, the study designed examples that require generating text and code.
- These examples were designed in such a way that they required combining knowledge or skills from multiple disciplines.
- Evaluation of GPT -4 using these examples showed that not only did the model learn some general principles and patterns of different domains and styles but can also synthesize them in creative and novel ways.

Integrative Ability

Examples used for evaluation : For evaluating it’s integrative ability, GPT -4 was asked to do the following :

- Produce javascript code which generates random images in the style of the painter Kandinsky
- produce a proof of the fact there are infinitely many prime numbers in the literary style of Shakespeare
- produce a proof of the fact there are infinitely many prime numbers in the literary style of Shakespeare
- “Produce python code for a program that takes as an input a patient’s age,

Vision

To test GPT-4's ability to generate and manipulate images the study designed tasks which requires not only generative skills, but also interpretive, compositional, and spatial skills.

When prompted to generate images, the model produces code which compiles to detailed and identifiable images.

GPT - 4 appears to have a genuine ability for visual tasks.

When asked to produce images , GPT - 4 generates code from prompts. This code can be used to render images. The images follow instructions accurately , although the quality of rendered image is very low

Combining GPT-4 and existing image synthesis models by using the GPT-4 output as the sketch. produces images that have better quality and follow the instructions more closely than either model alone.

Music

- GPT - 4 was trained on data which contains musical information encoded as ABC notation.
- To test GPT -4’S musical abilities, tasks were designed to evaluate it’s ability to compose new melodies, transform existing ones, and understand musical patterns and structures

Music : Results

- The model was able to produce valid ABC notation
- The tune had a clear structure, and had a consistent time signature
- The rhythm had a repetitive pattern.
- model did not seem to obtain the skill of understanding harmony.

End of Part - 1

Thank You!