



Auto-Wiki

Skim Smarter, Not Harder with AI-Powered Summaries



What for Who



Jef Packer

jef@sendit.ai



The Problem

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrmann
Eric Horvitz Edo Itamar Fler Peter Lee Yixuan Li Scott Lundberg
Harsha Nori Hamid Palangi Marro Tullio Ribeiro Yi Zhang
Microsoft Research

Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a vast range of learning and cognition. The latest model demonstrated a scale of cognitive and task ability that is the best of GPT-4, when it was still in active development by OpenAI, is part of a new cohort of LLMs (along with GPT-4o) more general intelligences than previous AI models. These models. We demonstrate that, beyond its ability to tackle tasks that span mathematics, coding, video, and special prompting. Moreover, in all of these tasks, performance and often nearly matches prior model. GPT-4 capabilities, we believe that it could reason about the world in a way that is more similar to human intelligence than any other AI system. In our exploration of GPT-4, we put special emphasis on discovering its limitations, and we discuss the challenges ahead for advancing towards deeper and more comprehensive versions of AGI, including the possible need for pursuing a new paradigm that moves beyond text-based prediction. We conclude with reflections on societal influences of the recent technological leap and future research directions.

1 / 155

Approaches

- Boosted Prompt Ensembles for Large Language Models (April 2023)
- Global Prompt Cell: A Portable Control Module for Effective Prompt (April 2023)
- Why think step-by-step? Reasoning emerges from the locality of experience (April 2023)
- Revisiting Automated Prompting: Are We Actually Doing Better? (April 2023)
- REFINER: Reasoning Feedback on Intermediate Representations (April 2023)
- Reflexion: an autonomous agent with dynamic memory and self-reflection (March 2023)
- CoVE: Communicative Agents for "Mind" Exploration of Large Scale Language Model Society (Mar 2023)
- Self-Refine: Iterative Refinement with Self-Feedback (Mar 2023)
- M3P: Multi-Prompting: Beyond-Context Learning with Calibration-Free Nearest Neighbor Inference (Mar 2023)
- Visual-Language Prompt Tuning with Knowledge-guided Context Optimization (Mar 2023)
- Fairness-guided Few-shot Prompting for Large Language Models (Mar 2023)
- Context-faithful Prompting for Large Language Models (Mar 2023)
- Is Prompt AI You Need? No, A Comprehensive and Broader View of Instruction Learning (Mar 2023)
- UPRISE: Universal Prompt Retrieval for Improving Zero-Shot Evaluation (Mar 2023)
- Model-tuning Via Prompts Makes NLP Models Adversarially Robust (Mar 2023)
- Structure Pretraining and Prompt Tuning for Knowledge Graph Transfer (March 2023)
- CoTEVer: Chain of Thought Prompting Annotation Toolkit for Explanation Verification (March 2023)
- Larger language models do in-context learning differently (March 2023)
- OpenCL: An Open-Source Framework for In-context Learning (March 2023)
- Dynamic Prompting: A Unified Framework for Prompt Tuning (March 2023)

Papers

The following are the latest papers (sorted by release date) on prompt engineering. We update this on a daily basis and new papers come in. We incorporate summaries of these papers to the guides above every week.

Overviews

- Text Learning with Foundation Models (April 2023)
- Deep Seat Strap for Generative AI, One-Click Leap for AGI: A Complete Survey on ChatGPT in AGC Era (April 2023)
- A Bibliometric Review of Large Language Models Research from 2017 to 2023 (April 2023)
- A Survey of Large Language Models (April 2023)
- Nature Language Reasoning, A Survey (Mar 2023)
- Augmented Language Models: A Survey (Feb 2023)
- A Survey for In-context Learning (Dec 2022)
- Towards Reasoning in Large Language Models: A Survey (Dec 2022)
- Reasoning with Language Model Prompting: A Survey (Dec 2022)
- Emergent Abilities of Large Language Models (Jun 2022)
- A Taxonomy of Prompt Modifiers for Text-to-Image Generation (Apr 2022)
- Prompt, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing (Jan 2022)

Other News + Tools

- A graph dialogue with LLMs will become the norm in the future and visualise thought processes [Link^[442]]. Work is being done source and available to the public.
- Replit have an interesting article on how they train LLMs. They some of their models [Link^[443]].
- If you're wondering how search might look with chatgpt, Multi-task that showcases what it will look like [Link^[444]]. It even manage [Link^[445]].
- A web ui of autogpt on huggingface [Link^[446]].
- Brevex becomes one of the first companies to actually use AI as they used image tools like Controlnet to create brand images [Link^[447]].
- An AI playground similar to *nat.dev*^[392] by Vercel. Use this to create and their outputs [Link^[448]].
- Someone connected ChatGPT to their personal health data and their health. This will be massive in the future. Generously surprised raise 50M+ VC money to transform digital health with open source [Link^[449]].
- Midjourney is releasing tutorials on how to get started coding with learn, this is free and a good starting point - a simple Q&A bot youtube video [Link^[450]]. Link to Replit [Link^[451]]. If you don't become familiar with it, its good.
- Reddit will begin same with Stack Overflow [Link^[452]].
- Microsoft has been to train LLMs like GPT-4. It seems like the old on code. Unfortun article [Link^[453]].
- Cheng is interested that knows what some of the Scale AI released budget but over 10 healthcare, financial assessment, ops etc [Link^[454]].
- An interesting thread [Link^[455]].
- ChatGPT talking about [Link^[456]].
- Malaysian announced BerkeleyQuest - a Berkeley [Link^[457]].

Text-to-video

- NVIDIA released their text-to-video getting better so fast, its going generation now. Being able to when you consider what some of Adobe released their text-to-video generate sound effects/music [Link^[458]].

AR + AI

- AR + AI for cooking, looks cool [Link^[393]].
- AR + AI for 3D knowledge mapping, looks so cool. If you have a metaquestvr you can download and try it [Link^[377]].

Law

- Two comedians made an AI tom brady say funny stuff. He threatened to sue. This is going to be very common going forward [Link^[394]].
- A german magazine did an "interview" with an AI Michael Schumacher and his family is now gone sue them [Link^[395]].
- An AI copilot for lawyers [Link^[403]].
- A lawyer discusses how he uses ChatGPT daily, an interesting thread [Link^[411]].

Finance

- Fintech is chatgpt for finance - ask questions about public companies. It provides reasoning, sources and data [Link^[423]].

Elon's TruthGPT

- Elon Musk went on Tucker Carlson and spoke about AI. He's building his own AI called TruthGPT - a maximum truth-seeking AI that tries to understand the nature of the universe. Whatever that means. This comes only a few weeks after he called for a pause in AI advancements. Why's he doing this? He was scared that Google/DeepMind were winning and would lead to unsafe AGI because Larry Page (co-founder of Google) called Elon a "species-ist" for being pro human because he wants AI to be safe for humanity. Page has openly stated that Google's goal is to create AGI [Link^[192]].

OpenAI TED Talk

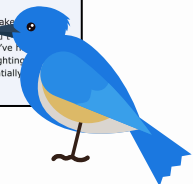
- President and Co-Founder of OpenAI, Greg Brockman did a TED talk and it's worth a watch. He showcased the potential for plugins in chatgpt and ends with "We all need to become literate, together I believe we can achieve the OpenAI mission of ensuring AGI benefits all of humanity". Another interesting point is that chatgpt or plugins is essentially "a unified language interface on top of tools". Genuinely wonder what they have access to behind the scenes [Link^[420]] [Link^[421]].

Games

- AI in Game dev - You can now connect any hugging face model in Unity. Open source API integration [Link^[422]]. This concept shows working AI in a game [Link^[423]]. Video showing how to connect the api [Link^[424]].
- A demo of using ChatGPT NPC's in virtual reality [Link^[425]].
- Someone made a game where you guess if the power of a lady is real or AI. I got 13/17 lol [Link^[426]]. A good way to show someone the power of AI but also highlights just how used to were seeing fake looking pics on social media
- AI powered 3D editor, looks cool [Link^[427]].

Music

- The music industry is about to undergo crazy change with AI songs of Drake, Kendrick Lamar, and others popping up and they are getting very good [Link^[428]]. Kanye, Drake singing Call Me Maybe & kpop is one of the funniest thing I've ever while lol [Link^[393]] [Link^[422]]. Obviously music companies are fighting this very hard. Will be very interesting how this plays out for artists essentially their voices as models to be bought or something like that [Link^[435]].



GPT-4 Technical Report

OpenAI

Abstract

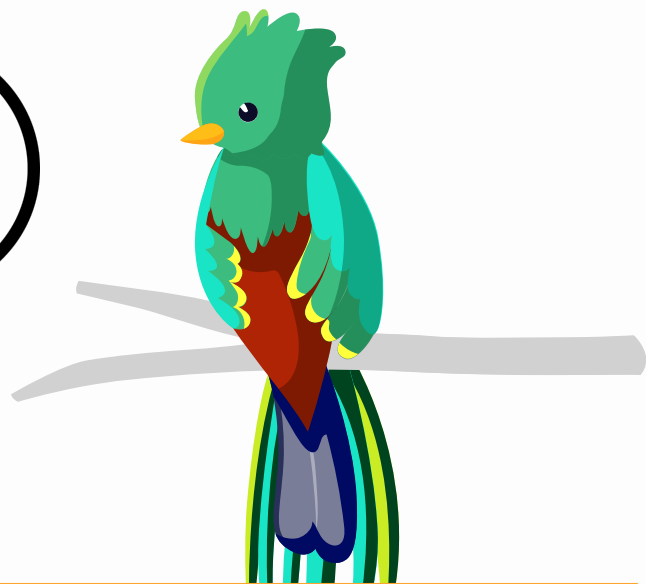
We report the development of GPT-4, a text-to-text model that can accept image and text inputs and produce human-like text. It performs well on various professional and academic benchmarks with a score around the top 10% of human performance. It is a text-to-text model pre-trained to predict the next token in a sequence. The alignment process results in improved adherence to desired behavior. A core component of this project was developing infrastructure and optimization methods that behave predictably across a wide range of scales. This allowed us to accurately predict some aspects of GPT-4's performance based on models trained with no more than 1/1,000th the compute of GPT-4.

1 / 100



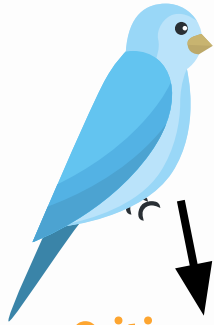
Solution

Auto-Wiki



How does it Work

Generate New Tasks

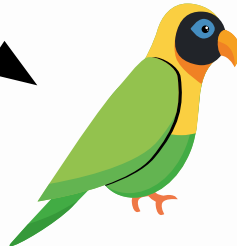


Critique Tasks



Tasks

Refine and Rerank Tasks



Memory

Complete Tasks



Tools



Read Doc
Write Doc
Google Search
...





Thank You!!