

# IMPACT OF AI ON JOURNALISM

Yogesh Haribhau Kulkarni



# Outline

① INTRODUCTION

② APPLICATIONS

③ DEMO

④ CONCLUSION

# About Me

# Yogesh Haribhau Kulkarni

## Bio:

- ▶ 20+ years in CAD/Engineering software development
- ▶ Got Bachelors, Masters and Doctoral degrees in Mechanical Engineering (specialization: Geometric Modeling Algorithms).
- ▶ Currently doing Coaching in fields such as Data Science, Artificial Intelligence Machine-Deep Learning (ML/DL) and Natural Language Processing (NLP).
- ▶ Feel free to follow me at:
  - ▶ Github ([github.com/yogeshhk](https://github.com/yogeshhk))
  - ▶ LinkedIn ([www.linkedin.com/in/yogeshkulkarni/](https://www.linkedin.com/in/yogeshkulkarni/))
  - ▶ Medium ([yogeshharibhaukulkarni.medium.com](https://yogeshharibhaukulkarni.medium.com))
  - ▶ Send email to [yogeshkulkarni at yahoo dot com](mailto:yogeshkulkarni@yahoo.com)

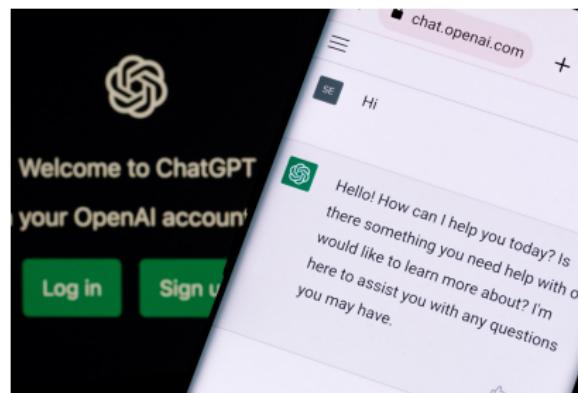


Office Hours:  
Saturdays, 2 to 5pm  
(IST); Free-Open to all;  
email for appointment.

# Overview of ChatGPT

# What is ChatGPT?

- ▶ A Chatbot
- ▶ A language model, that takes 'prompt' from user and generates a response.
- ▶ Built by OpenAI and released in Nov 2022
- ▶ Got 1m users in 5 days (Insta 2.5 months, Spotify 5m, Facebook 10m, Netflix 3.5yrs)
- ▶ Link: <https://chat.openai.com/chat>
- ▶ Details: <https://openai.com/blog/chatgpt/>

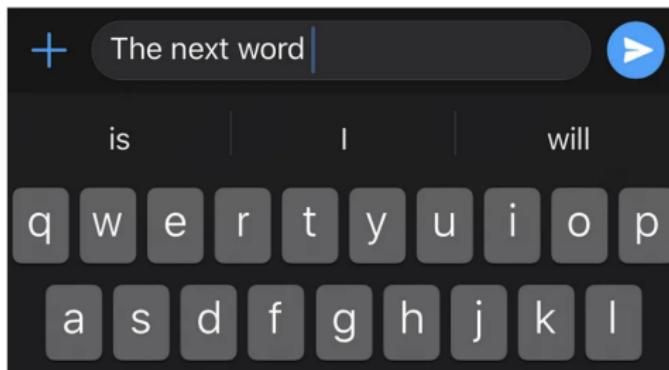


(Ref: OpenAI, creator of ChatGPT, casts spell on Microsoft - The Jakarta Post

YHK

## What is a Language Models?

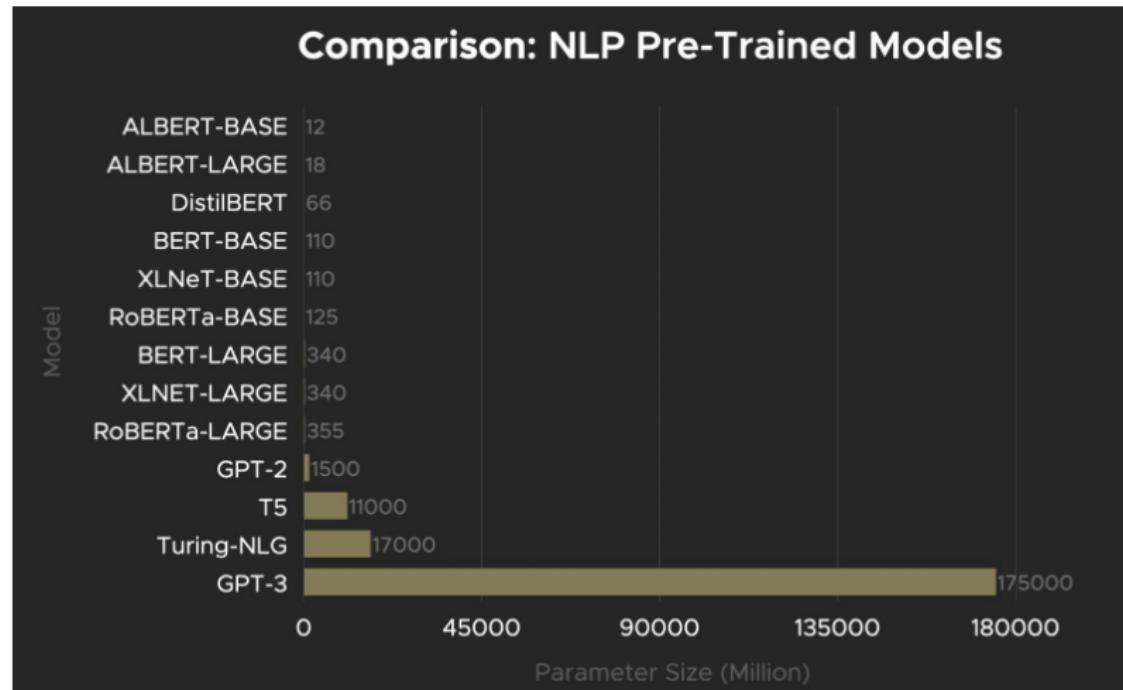
- ▶ While typing SMS, have you seen it suggests next word?
- ▶ While typing email, have you seen next few words are suggested?
- ▶ How does it suggest? (suggestions are not random, right?)
- ▶ In the past, for "Lets go for a ... ", if you have typed 'coffee' 15 times, 'movie' say 4 times, then it learns that. Machine/Statistical Learning.
- ▶ Next time, when you type "Lets go for a ", what will be suggested? why?
- ▶ This is called Language Model. Predicting the next word. When done continuously, one after other, it spits sentence, called Generative Model.



Next word prediction using language modeling in keyboards(Mandar Deshpande)

YHK

# Large Language Models - Comparison



(Ref: Deus.ai <https://www.deus.ai/post/gpt-3-what-is-all-the-excitement-about>)

# Open AI: Who are these guys?

- ▶ San Francisco-based artificial intelligence company
- ▶ Famous for its well-known DALL-E, generates images from prompts.
- ▶ Initially supported by Elon Musk
- ▶ The CEO is Sam Altman, who previously was president of Y Combinator.
- ▶ Microsoft is a partner and investor.
- ▶ Mission: To ensure that artificial general intelligence benefits all of humanity. Prevent misuse of AI

## OpenAI Founded

The company was founded with the goal of developing and promoting friendly AI in a responsible way, with a focus on transparency and open research.



(Ref: <https://www.slideegg.com/open-ai-chat-gpt>)

# GPTs Training

## GPT: Generative Pre-trained Transformers

- ▶ GPT-1 is pre-trained on the BooksCorpus dataset, containing 7000 books amounting to 5GB of data
- ▶ GPT-2 is pre-trained using the WebText dataset which is a more diverse set of internet data containing 8M documents for about 40 GB of data
- ▶ GPT-3 uses an expanded version of the WebText dataset, two internet-based books corpora that are not disclosed and the English-language Wikipedia which constituted 600 GB of data

## GPTs Training compared to human reading

- ▶ GPT-3 was trained on 499B tokens; GPT-4, on 1.4T tokens.
- ▶ In comparison, if you spent 12 hours a day reading for an entire lifetime (80 years) at average speed (250 words / minute), we would absorb 5.26B words (tokens).
- ▶ That's a ratio of 100:1 between the training data used for GPT-3 and the amount of data that can ever be read by a human, and 260:1 for GPT-4.

(Ref: LinkedIn post by Dr Jennifer Prendki)

# GPT3 vs ChatGPT

	GPT-3	ChatGPT
<b>Training Parameters</b>	175 billion	175 Billion RLHF Technique Human conversational text
<b>Designed For</b>	NLP Tasks	Chatbot apps
<b>Content Retention Ability</b>	No	Yes
<b>Use Cases</b>	Creative Writing, Business Writing, Ads, Translation, Summarization, Data Analysis	Customer Service, Data Analysis, Blogging, Copywriting, Coding, Debugging, etc.

(Ref: ChatGPT Explained: Complete A-Z Guide - Kripesh Adwani)

# Applications

# Explain Complex Subjects



What is neuroscience?



Neuroscience is the scientific study of the nervous system, which includes the brain, spinal cord, and peripheral nerves. It is a multidisciplinary field that combines knowledge from biology, chemistry, physics, psychology, and other fields to understand how the nervous system works and how it relates to behavior and cognition. Researchers in neuroscience use a variety of methods, including neuroanatomy, neurophysiology, and neuroimaging, to study the structure and function of the nervous system.



(Ref: Top 10 Chat GPT Use Cases - Simplilearn)

# Write Code

The screenshot shows a GPT-3 AI interface. A user prompt is at the top: "Give me a machine learning sample program for stocks prediction". Below it, the AI responds with a Python code snippet. The code imports numpy and scikit-learn's LinearRegression and train\_test\_split modules. It then loads stock price data from a file named 'stock\_prices.txt'. A red arrow points to the word 'LinearRegression' in the code. On the right side of the code block, there is a "Copy code" button.

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

# Load stock data
stock_prices = np.genfromtxt('
```

it was still writing but I stopped it, just to show this pic

(Ref: Top 10 Chat GPT Use Cases - Simplilearn)

# Debug Code

```
arr[j + 1] = tempp;
```

It should be:

 Copy code

```
arr[j + 1] = temp;
```

You have an extra "p" in the variable name, ~~temp~~ which is causing a compile error. This is because "temp" is the variable that you declared earlier to store the value of the current element, whereas "tempp" is not a variable that has been declared.

This error is causing the code not to run. Once you fix this issue, the code should work correctly and sort the array in ascending

(Ref: Top 10 Chat GPT Use Cases - Simplilearn)

# Get Custom Marketing Strategy



How should i promote my youtube videos on social media?



Promoting your YouTube videos on social media can be a great way to increase views and engagement. Here are a few strategies you can use:

1. Share your videos on your personal social media accounts. This includes your Facebook, Twitter, Instagram, and other accounts where you have a following. Encourage your friends, family and followers to share the video with their own networks.
2. Join relevant communities and groups on social media. For example, if you're making cooking videos, join Facebook groups and Reddit communities dedicated to cooking. Share your videos with the group and engage with the community.
3. Use social media ads. Platforms like Facebook, Instagram, and Twitter allow you to create targeted ads that promote your videos



## Write Articles



Write an article on machine learning in 500 words



Machine learning is a rapidly growing field that is transforming the way we live and work. It is a branch of artificial intelligence that allows computers to learn from data, without being explicitly programmed. Machine learning is changing the way we interact with technology and has the potential to revolutionize many industries, including healthcare, finance, transportation, and more.

The basic idea behind machine learning is that computers can learn from data, without being explicitly programmed. This is done by feeding a computer large amounts of data and allowing it to find patterns and make predictions. There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is the most common type of machine learning. It involves training a computer on a labeled dataset, where the desired output is ■

## Summarize Book



summarize the book Gullivers travels



Gulliver's Travels is a satirical novel written by Jonathan Swift in 1726. The book tells the story of Lemuel Gulliver, a ship's surgeon who becomes shipwrecked and ends up on a series of strange and fantastical islands. The novel is divided into four parts, each of which tells the story of Gulliver's travels to a different island.

In Part 1, Gulliver washes up on the island of Lilliput, where he encounters a race of tiny people who are locked in a bitter war with their neighboring island of Blefuscus. Gulliver becomes embroiled in the politics of the island and ultimately helps the Lilliputians defeat the Blefuscudians.

In Part 2, Gulliver travels to the island of Brobdingnag, where he encounters a race of giants. He is initially treated kindly by ■

# Answer Interview Questions



How to answer the interviewer what is your weakness



When answering this question, it is best to be honest, but also to frame your answer in a positive way. For example, you could say something like: "I have a tendency to be a perfectionist and sometimes struggle with delegating tasks. However, I have been working on this by learning to trust my team members and giving them more responsibility." This shows that you are aware of your weakness and actively working to improve it.

(Ref: Top 10 Chat GPT Use Cases - Simplilearn)

## Develop Apps



I want to write an android app that shows me 50 stock prices as a list from yahoo API. How can i do it?



To write an Android app that shows a list of 50 stock prices from the Yahoo API, you will need to do the following:

1. Create a new Android project using Android Studio or another development environment.
2. Register for a Yahoo API key, which will allow you to access financial data such as stock prices.
3. Use the Retrofit library to make API calls to the Yahoo API. You can use the Retrofit library to create an interface that defines the API endpoints, and then use this interface to make API calls.

4 ■

## Create Health Plan



My weight is 95 kg and I want to loose it. create 7 days plan excercise and diet planfor it.



Here is a 7-day exercise and diet plan that can help you lose weight:

Day 1:

- Exercise: 30-minute brisk walk or jog
- Diet: Start your day with a healthy breakfast of oatmeal, berries, and a boiled egg. For lunch, have a salad with grilled chicken or fish. For dinner, have grilled vegetables and a small serving of lean protein.

Day 2:

- Exercise: 45-minute circuit training routine that includes bodyweight exercises such as push-ups, ■

# Answer General Knowledge Questions



Who invented the first printer?



The first printer was invented by a German goldsmith named Johannes Gutenberg in the 15th century. He developed a printing press that used movable type, which made it possible to print books and other documents quickly and cheaply. The Gutenberg press used metal movable type, which was set into a wooden form and inked, then pressed onto paper. This invention revolutionized the way information was disseminated and made the mass production of books possible.

Gutenberg's invention of the printing press is widely regarded as one of the most important events in the second millennium, it is considered the most important invention of the second millennium, after the invention of the wheel.

(Ref: Top 10 Chat GPT Use Cases - Simplilearn)

# Applications of Generative AI in Journalism

## Introduction to AI in Journalism

- ▶ AI encompasses a range of technologies that simulate human intelligence.
- ▶ Current capabilities include natural language processing (NLP), image recognition, and content generation.
- ▶ These capabilities are transforming journalism by automating tasks and enhancing efficiency.
- ▶ AI algorithms can analyze vast amounts of data quickly, leading to insights and story leads.
- ▶ Journalists are increasingly relying on AI tools to streamline their workflow and produce high-quality content.

# AI Applications in Journalism

- ▶ Automated News Writing: From structured data, such as financial reports or sports scores. Also in producing routine news stories quickly, freeing up journalists for more in-depth reporting.
- ▶ Content Personalization (electronic media): based on individual preferences and interests.
- ▶ Fact-Checking: Cross-referencing multiple sources and databases.
- ▶ Audience Analytics: insights into audience preferences, helping journalists create content that resonates with their readers/viewers.

## Advantages of AI in Journalism

- ▶ Faster and More Efficient News Production: automates routine tasks like news writing, enabling journalists to focus on more complex stories.
- ▶ Ability to Analyze and Process Large Volumes of Data: to uncover patterns and insights that may be overlooked by human journalists.
- ▶ Personalized Content Delivery Based on User Preferences: delivering relevant stories and updates, increasing audience satisfaction and loyalty.

## Challenges and Concerns

- ▶ Potential Job Displacement and Automation of Journalistic Tasks:  
Journalists may need to acquire new skills to adapt to the changing landscape and work alongside AI systems.
- ▶ Bias and Ethical Concerns in AI-Generated Content:
  - ▶ May inadvertently perpetuate biases present in training data, leading to skewed or unfair news coverage.
  - ▶ Ensuring fairness and impartiality in AI-generated content requires ongoing oversight and transparency.
- ▶ Credibility and Trust Issues with AI-Written Articles:
  - ▶ Readers may question the credibility of news articles generated by AI, raising concerns about transparency and authenticity.
  - ▶ Media organizations need to disclose the use of AI in content creation and uphold editorial standards to maintain trust with their audience.
- ▶ Privacy and Data Protection Concerns: Who OWNs the content, even though it is publicly available? Who is making money out of it? Who will take up legal cases for misinformation or defamation?

## Case Studies and Examples

- ▶ **The Washington Post's Heliograf:** to produce automated news stories on topics like sports, elections, and weather, allowing journalists to focus on more complex reporting.
- ▶ **Reuters' Lynx Insights:** to analyze large datasets and identify news trends, to discover story leads and produce data-driven content across various beats.
- ▶ **The Associated Press' AP Insights:** to personalize news content for its subscribers.
- ▶ **BBC's Project Juicer:** to analyze audience engagement and optimize content distribution.

## The Role of Human Journalists

- ▶ Importance of Human Oversight, Editing, and Fact-Checking: to provide oversight and quality control, verifying information, detecting biases, and correcting errors in AI-generated content.
- ▶ Unique Skills and Qualities that Human Journalists Bring:
  - ▶ Critical thinking, intuition, and empathy, enabling them to contextualize news events and interpret complex issues for their audience.
  - ▶ Storytelling prowess, investigative skills, and subject matter expertise that AI algorithms cannot replicate.
- ▶ Collaboration Between AI and Human Journalists: Complementary by automating routine tasks, analyzing data, and uncovering insights, allowing journalists to focus on higher-order tasks. Human with AI skills is more dangerous than just AI.

## Ethical Considerations and Guidelines

- ▶ Transparency and Disclosure: the use of AI in content creation and clearly label AI-generated content as such.
- ▶ Fairness and Bias Mitigation:
  - ▶ Journalists must be vigilant in identifying and addressing biases in AI algorithms and datasets to ensure fair and balanced news coverage.
  - ▶ Regular audits and diversity assessments can help mitigate bias and promote inclusivity in AI-driven journalism.
- ▶ Accountability: Moral, Social, Legal (and Financial)
- ▶ Privacy and Data Protection: Obtaining informed consent from users and adhering to data protection regulations help mitigate privacy risks and safeguard user information.

## Future Trends and Predictions

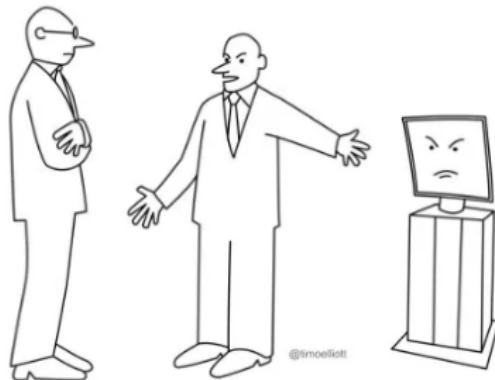
- ▶ Advancements in Natural Language Generation and Multimodal AI: capable of processing text, images, and audio together, will enable new forms of multimedia storytelling and news delivery.
- ▶ Potential for AI to Enhance Investigative Journalism and Data-Driven Reporting: to conduct deeper investigations and analyze large datasets more efficiently.
- ▶ Impact of AI on Media Business Models and Revenue Streams: to monetize content through subscription models and targeted advertising and will reduce operational costs and improve scalability, reshaping traditional media business models.

## Strategies for Adaptation and Upskilling

- ▶ Embrace Lifelong Learning:
  - ▶ You are domain (Journalism) expert. Need to know latest AI for usage in your domain. The Combination is Lethal, your niche.
  - ▶ Participating in workshops, online courses, and professional development programs can help journalists acquire new skills and knowledge.
- ▶ Develop Data Literacy and Analytical Skills: Training in data analysis, statistics, and data visualization will enable journalists to interpret and communicate insights from large datasets.
- ▶ Foster Collaboration and Interdisciplinary Teams: between journalists, data scientists, and AI experts to leverage diverse skills and perspectives.
- ▶ Cultivate Ethical Awareness and Critical Thinking: Training in media ethics, bias detection, and algorithmic transparency will help journalists navigate ethical dilemmas in the AI-driven future.

# Conclusion

## Opportunities



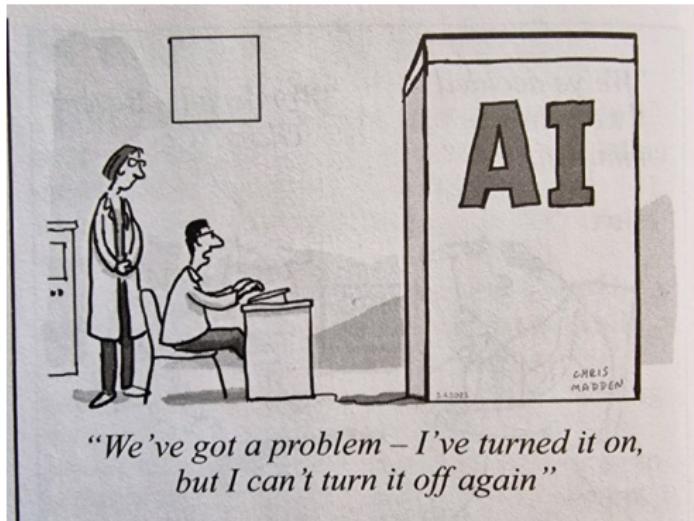
*His decisions aren't any better than yours  
— but they're WAY faster...*

- ▶ To streamline workflows, analyze data, and engage audiences in innovative ways.
- ▶ Automated content generation, personalized news delivery, and data-driven reporting

(ref: ProlitusTechnologies)

# Conclusion

## Challenges



(ref: Ian McVicar on LinkedIn)

- ▶ Ethical considerations, bias detection, and maintaining editorial standards
- ▶ Concerns about job displacement, privacy violations, and algorithmic accountability

## Lessons Learnt

- ▶ Embrace Lifelong Learning, Journalism + AI is the combination
- ▶ Human Supervision cannot be removed from the journalistic process, so no full automation
- ▶ AI models are as good as their training (data). Indic languages? non quantifiable data?
- ▶ Human generated content will be more premium (but has to be better than ChatGPT)

# Making of a Sandwich

# Basic

Explain how to make a peanut butter and jelly sandwich

## Adding Roles

As a chef, explain to your assistant how to make a peanut butter and jelly sandwich

## Adding Constraints

Make a nut-free version of the sandwich due to a customer's nut allergy

## Adding Examples

Create two unique variations of the classic sandwich. Banana Nut Crunch: . . .

Triple Berry Blast: . . .

## Adding Contextual Information

As the head chef at 'The Sandwich Haven,' guide your new assistant to create specials for the menu

## Incorporating Feedback

Improve the sandwich based on customer feedback for less sweetness and a creative twist

## Time Constraints and Prioritization

Prepare an alternative fruit version for testing within a tight deadline

## Incorporating Multidisciplinary Knowledge

Use food presentation and garnishing techniques for a visually appealing sandwich

## Addressing Dietary Preferences

Prepare a vegan version using plant-based alternatives for all ingredients

## Reflection and Iteration

Reflect on feedback and iteratively refine the sandwich for better taste and appeal

## Self-Criticism

Explain how to make a peanut butter and jelly sandwich. Please re-read your above response. Any mistakes? If so, please identify and make the necessary edits.

## Chain-of-Thought

Explain how to make a peanut butter and jelly sandwich. Let's think step by step.

## Self-Consistency

Here are recipes of multiple sandwiches. Sandwich 1: recipe 1. Sandwich 2: recipe 2. .... Explain how to make a peanut butter jelly sandwich.

# Conclusions

## Advantages

- ▶ Conversational Abilities
- ▶ Solving Complex Problems
- ▶ Retaining Previous Information
- ▶ Will replace mundane language tasks, How to articles, homework, etc

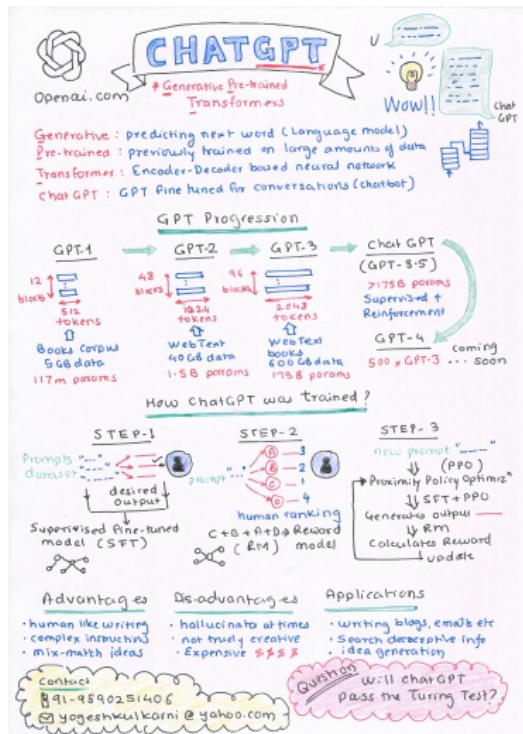
## Dis-advantages

- ▶ Sensitive to Input Phrasing
- ▶ Cannot replace humans for innovation, for which data does not exist already
- ▶ Keeps “hallucinating”: Tends to write plausible but incorrect content with confidence
- ▶ May not get language structure right all the time, e.g try getting ghazal written

## Compared to humans?

- ▶ Although GPT models are far more knowledgeable due to vast training (GPT-3 was trained on 499B tokens, GPT-4 on 1.4 T tokens whereas a human can read about 5.26B tokens in 80 years)
- ▶ It's not about the volume of data ...
- ▶ Children can talk long before they reach adulthood, and it's likely that < 1B tokens is enough for a human to master language. Need to figure out how nature has made her neural network so efficient.
- ▶ Imagine the pretraining done over billions of years at the time a baby comes ready with

# My Sketchnote



(Ref: <https://medium.com/technology-hits/overview-of-chatgpt-95f4b43645c0>)

# The After-shocks

## ChatGPT vs Google: Who is better?



Google: almost latest data, gives source, not trainable, looks more accurate

ChatGPT: a bit old data, no source, trainable, hallucinates

(Ref: ChatGPT Explained: Complete A-Z Guide - Kripesh Adwani)

# Will ChatGPT Kill Jobs?



Repetitive, boring and standard, language based jobs, for sure. Need to be more creative, experiential to stand against ChatGPT.

(Ref:ChatGPT Explained: Complete A-Z Guide - Kripesh Adwani)

## New Job Roles?

Prompt Engineer: Preparing input to AI effectively to get the desired answer.  
Will need to AI works in the background plus domain knowledge. Give context, examples etc to prime the model to give short specific answers than the usual page-long ones (davinci GPT3 in this case)

Playground Q&A Save View code Share ...

I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth, I will give you the answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Unknown".

Q: What is human life expectancy in the United States?  
A: Human life expectancy in the United States is 78 years.

Q: Who was president of the United States in 1955?  
A: Dwight D. Eisenhower was president of the United States in 1955.

Q: Which party did he belong to?  
A: He belonged to the Republican Party.

Q: What is the square root of banana?  
A: Unknown

Q: How does a telescope work?  
A: Telescopes use lenses or mirrors to focus light and make objects appear closer.

Q: Where were the 1992 Olympics held?  
A: The 1992 Olympics were held in Barcelona, Spain.

Mode: text-davinci-003  
Temperature: 0  
Maximum length: 100  
Stop sequences: Enter sequence and press Tab  
Top P: 1  
Frequency penalty: 0  
Presence penalty: 0

(Ref: Advanced ChatGPT Guide - How to build your own Chat GPT Site - Drian Twarog)

# The Hype



Elon Musk ✅ @elonmusk · Dec 4

Replies to @sama

ChatGPT is scary good. We are not far from dangerously strong AI.

...



François Chollet ✅ @fchollet · 8h

With apps like Midjourney, Lensa, and ChatGPT, the age of consumer AI has finally arrived.

...



@jh@sigmoid.social (Mastodon) @jeremyphoward · 9h

...

I remember well when Google was first released. I felt confident that it would become a critical tool that we'd all rely on.

I haven't felt that way again, until now: ChatGPT.

In just a few days I've gotten so much out of it.

(Ref: ChatGPT - Intro & Potential Impact Sudalai Rajkumar, SRK)

## Finally, from horses mouth!!



Sam Altman  @sama · 10 dic.

ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness.

it's a mistake to be relying on it for anything important right now. it's a preview of progress; we have lots of work to do on robustness and truthfulness.

877 3.856 27,5 mil

(Ref: ChatGPT: training process, advantages, and limitations - By Sergio Soage, Machine Learning Engineer at Aivo)

## References

- ▶ Let's build GPT: from scratch, in code, spelled out: Andrej Karpathy
- ▶ ChatGPT and Reinforcement Learning - CodeEmporium

# Thanks ...

- ▶ Search "**Yogesh Haribhau Kulkarni**" on Google and follow me on LinkedIn and Medium
- ▶ Office Hours: Saturdays, 2 to 5pm (IST); Free-Open to all; email for appointment.
- ▶ Email: yogeshkulkarni at yahoo dot com

(<https://www.linkedin.com/in/yogeshkulkarni/>, QR by Hugging Face

QR-code-AI-art-generator, with prompt as "Follow me")

