# From Zero to ChatGPT!!

Yogesh Haribhau Kulkarni

YHK

# Outline

# About Me

YHK

# 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)
  - LinkedIn (www.linkedin.com/in/yogeshkulkarni/)
  - Medium (yogeshharibhaukulkarni.medium.com)
  - Send email to yogeshkulkarni at yahoo dot com

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

YHK

# From Zero to GPT!!

YHK

# Chess: next move?

- Needs extreme expertise
- Needs "intelligence"
- How do you get that?
  - Built by lots of training.
  - By studying lots of past games.
- This is how Humans build intelligence

YHK

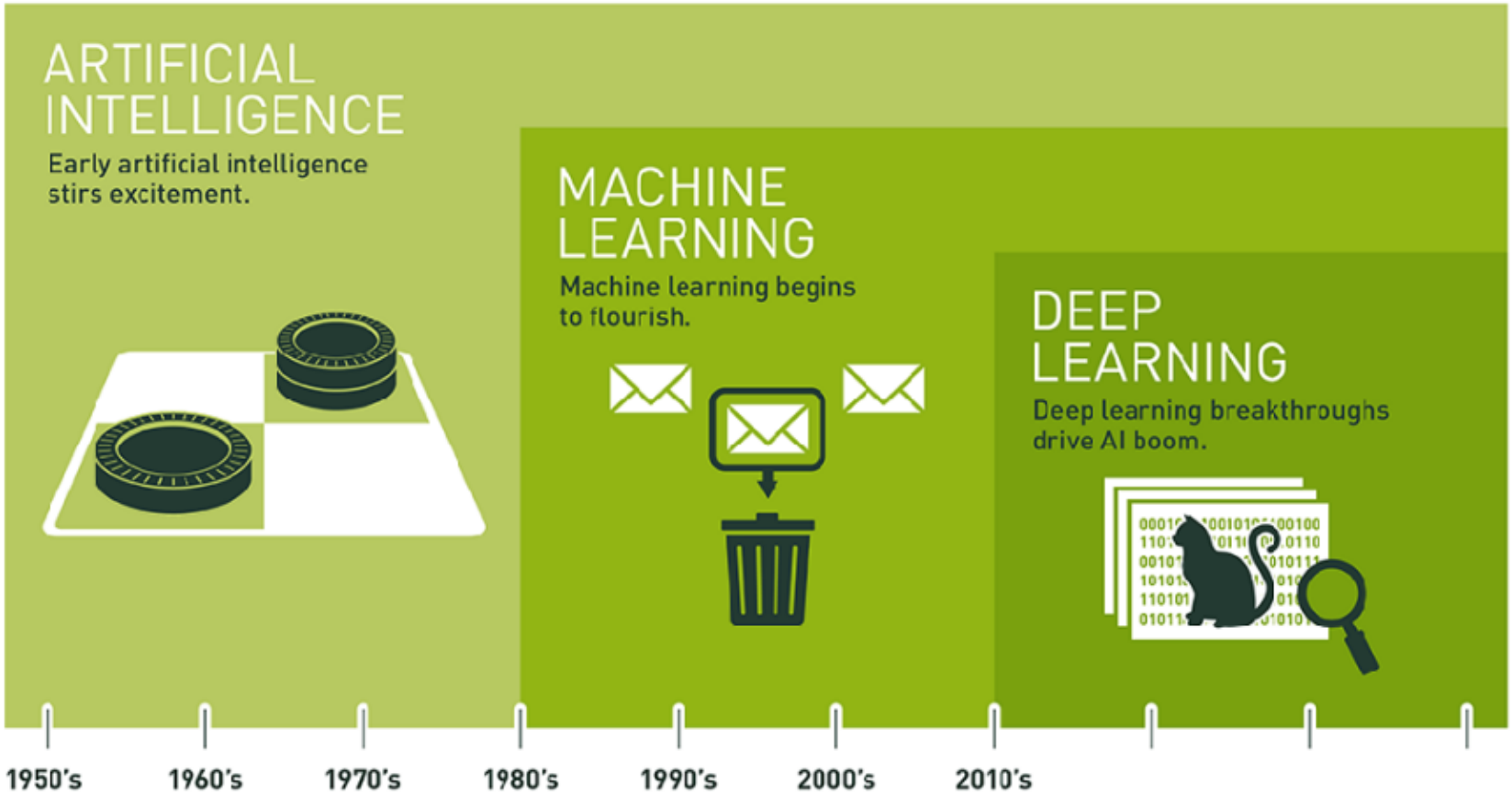
# Intelligence

- Can machine (software/program) also do the same?
- Can it play chess?
- Can it build intelligence?
- By looking at past experiences (data),
- Training Data: games played, moves used, etc.

Yes, it can!! Thats Artificial Intelligence.

YHK

# Relationship between AI, ML, DL




(Ref: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/)

YHK

# What is AI-ML-DL?

- Artificial Intelligence: mimicking human intelligence
- Machine Learning: Automating Learning with features.
- ML: human-designed representations and input features. So, its just optimizing weights to best make a final prediction
- There could be programmed (hand coded) AI, that's not Machine Learning
- Machine Learning could be for non AI activities, like automation
- Deep Learning: Neural network with no input features

YHK

# Introduction to Machine Learning

YHK

# How do we learn?

- What do we do when we have to prepare for an examination?
- Study. Learn. Imbibe. Take notes. Practice mock papers.
- Thus, prepare for the unseen test.
- Machine Learning does the same.

YHK

# What is Machine Learning?

Machine learning is a type of artificial intelligence (AI) which:

- Learns function without being explicitly programmed.
- Can grow and change when exposed to new data.

YHK

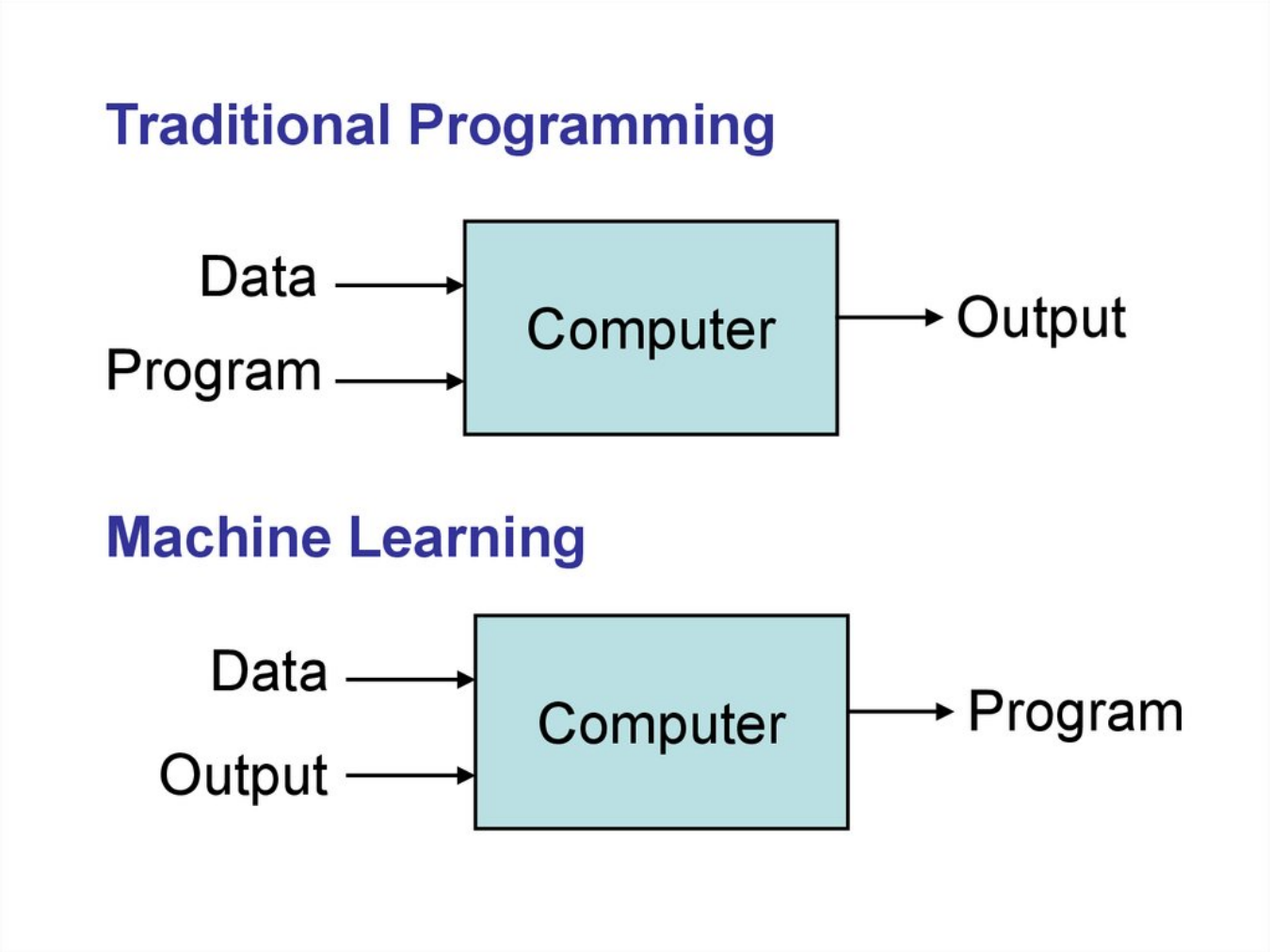
# Model entities

For $income = c + \beta_0 \times education + \beta_1 \times experience$

- Inputs: Education and experience, also called as features or attributes or dimensions or variables.
- Mathematical entities added to input data, are Parameters.. $\beta_0$ and $\beta_1$ are parameters
- Income is target, also called as outcome or class.

YHK

# Traditional vs. Machine Learning?



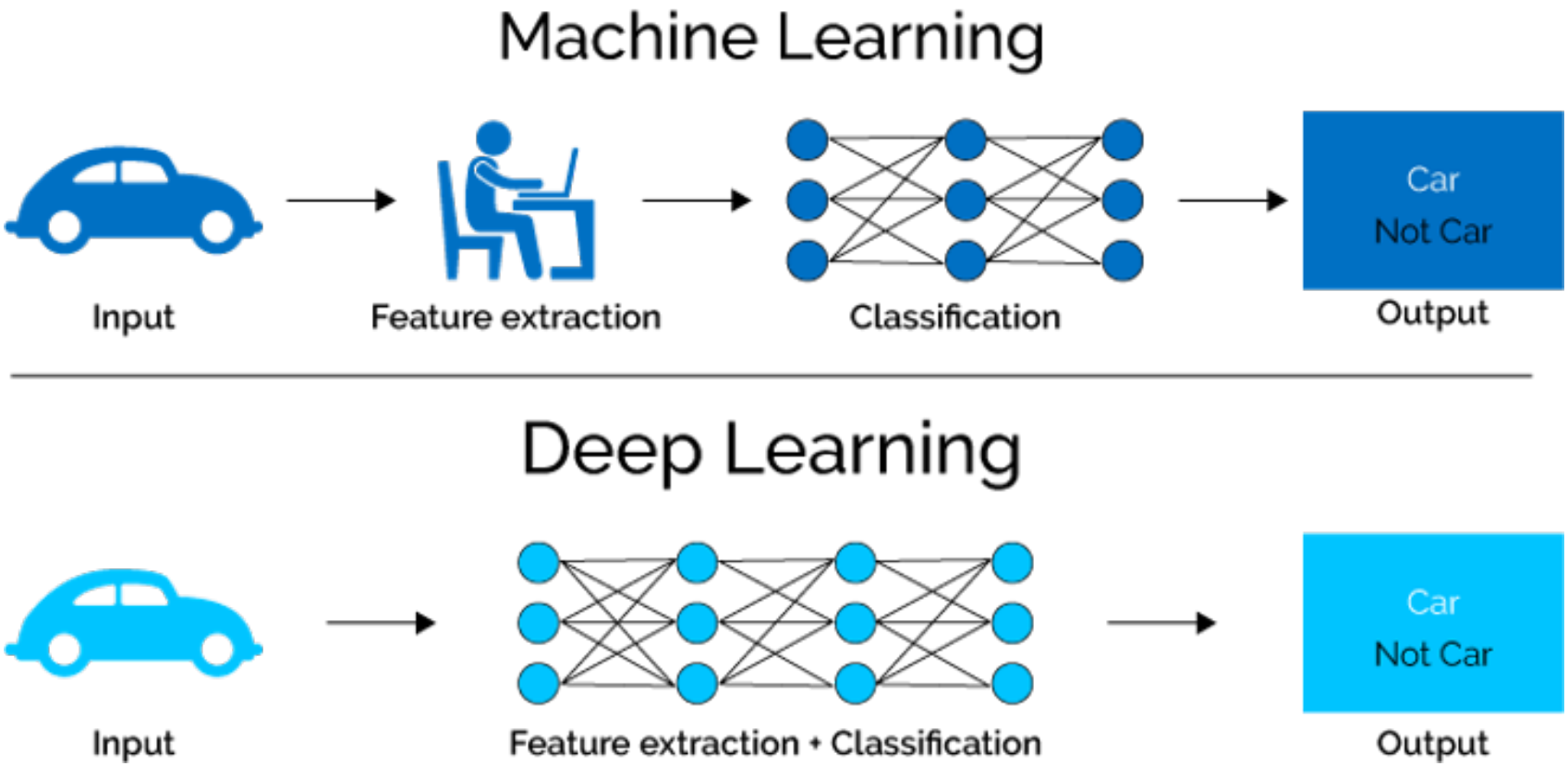
# Why Machine Learning?

- Problems with High Dimensionality
- Hard/Expensive to program manually
- Techniques to model ‘ANY’ function given ‘ENOUGH’ data.
- Job $$$

YHK

# ML vs DL: What's the difference?

Deep learning algorithms attempt to learn (multiple levels of) representation by using a hierarchy of multiple layers




(Reference: https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png)

YHK

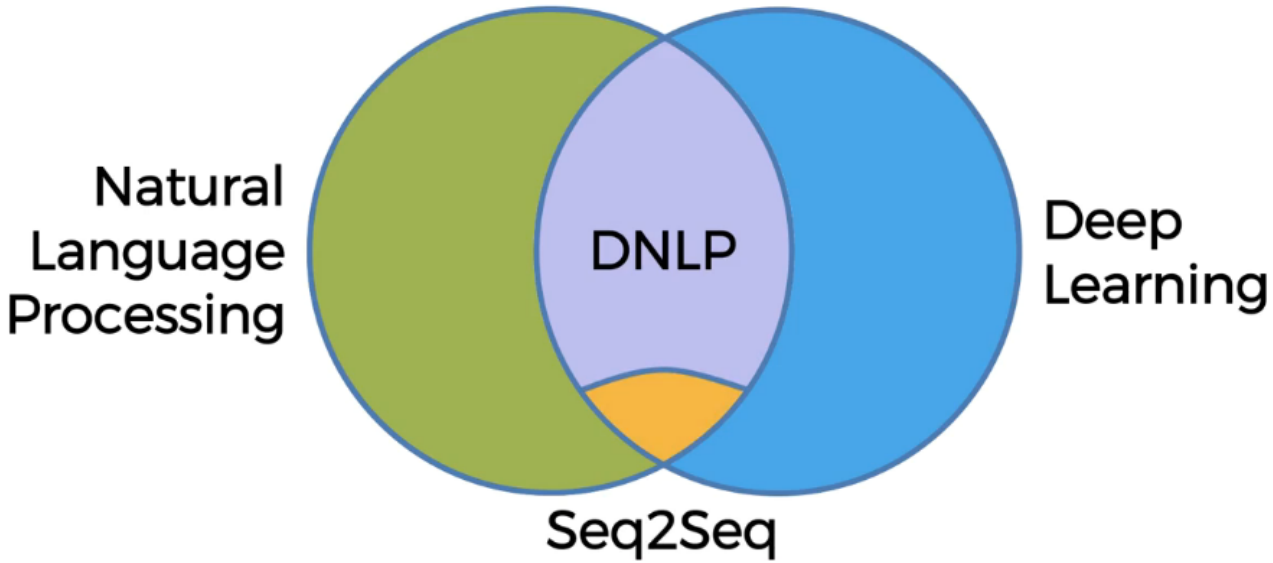
# Use Deep Learning When ...

- You have lots of data (about 10k+ examples)
- The problem is "complex" - speech, vision, natural language
- The data is unstructured
- You need the absolute "best" model

(Ref: Introduction to TensorFlow 2.0 - Brad Miro)

YHK

# What is Deep NLP



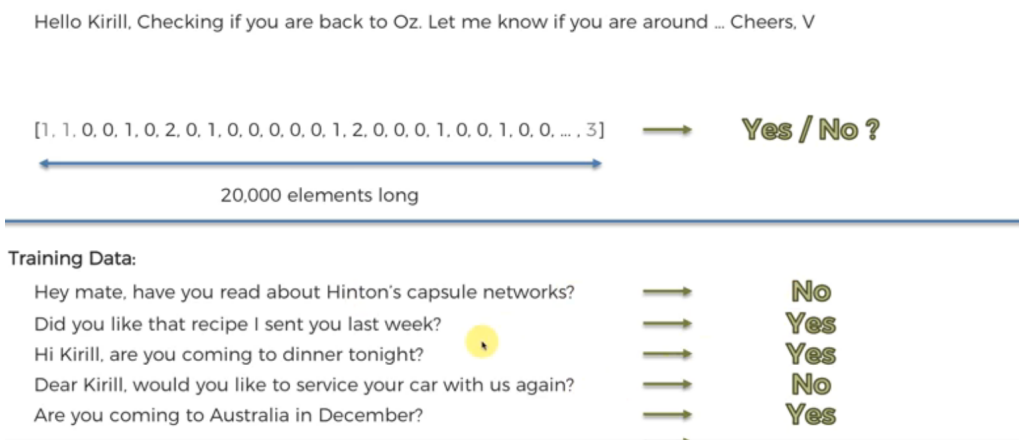

(Ref: Deep Learning and NLP A-Z - Kirill Eremenko)
(Note: Size is not indicative of importance)

- Green part is NLP (rule based, linguistic)
- Blue part is Deep Learning not applied to NLP
- Purple is Deep NLP (DNLP), NN applied for NLP use cases
- Seq2Seq is heavily used technique of DNLP for sequence to sequence modeling, eg Translation, Q & A, etc.

YHK

# Typical Machine Learning Classification

- Questions are converted to bag of words (a vocab long vector, having frequency of specific words at their places)
- Each question thus gets converted to fixed size vector, which acts as list of features.
- In training, weights are computed based on the given target.
- Once model is ready, it is able to answer Yes or No to the question.




(Ref: Deep Learning and NLP A-Z - Kirill Eremenko)

YHK

# Context

Representing words by their context

- Distributional semantics: **A word's meaning is given by the words that frequently appear close-by**
  - *"You shall know a word by the company it keeps"* (J. R. Firth 1957)
  - One of the most successful ideas of modern statistical NLP!
- When a word *w* appears in a text, its **context** is the set of words that appear nearby (within a fixed-size window).
- Use the many contexts of *w* to build up a representation of *w*

| | | |
|---|---|---|
| *…government debt problems turning into* | ***banking*** | *crises as happened in 2009…* |
| *…saying that Europe needs unified* | ***banking*** | *regulation to replace the hodgepodge…* |
| *…India has just given its* | ***banking*** | *system a shot in the arm…* |

These context words will represent ***banking***

YHK

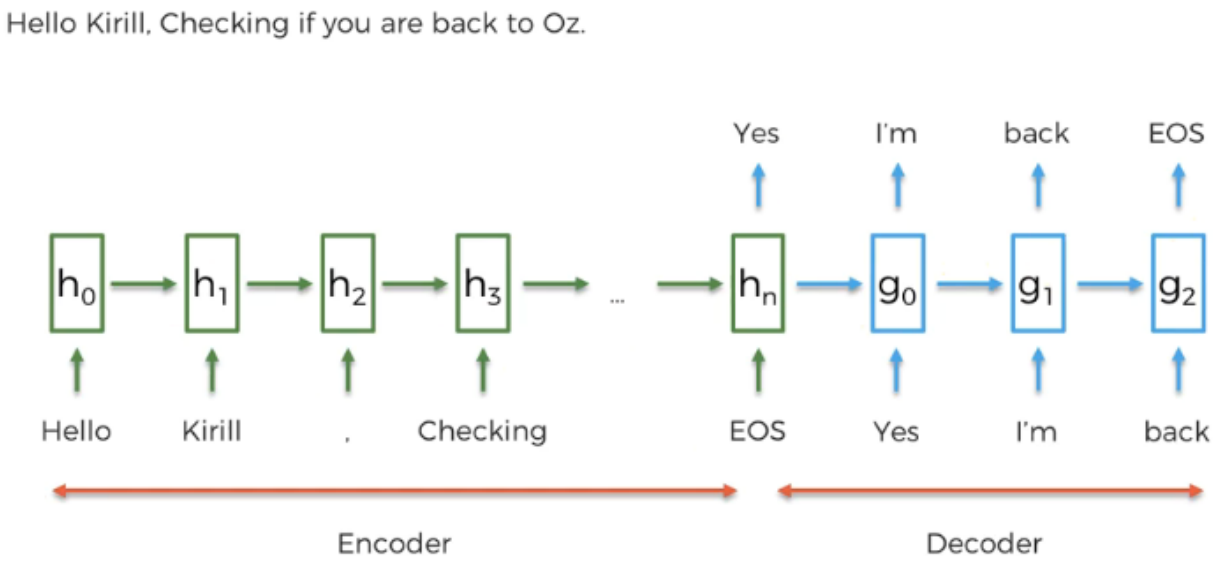
# Word vectors

- Dense vector for each word
- Called distributed representation, word embeddings or word representations
- Test: similar to vectors of words that appear in similar contexts

$$\textit{banking} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}$$

YHK

# Seq2Seq architecture



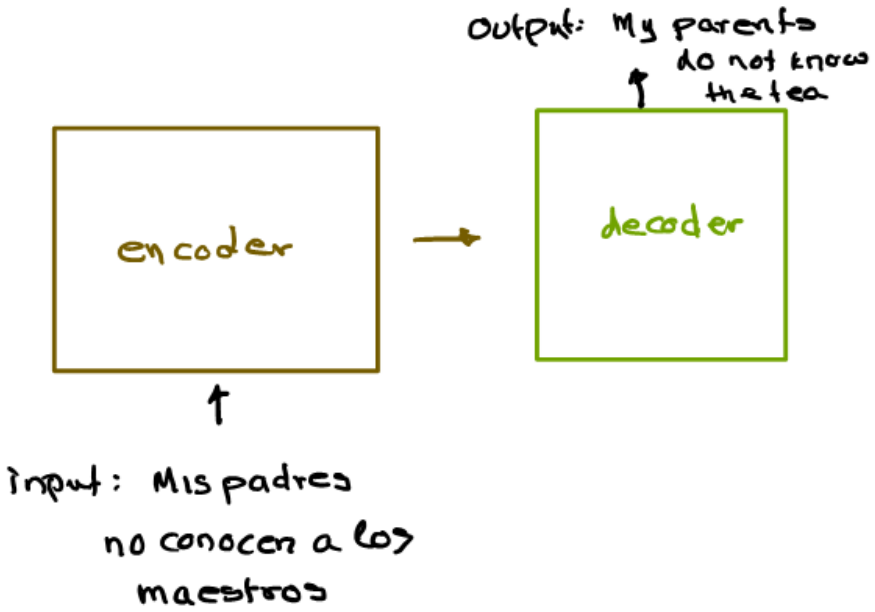

(Ref: Deep Learning and NLP A-Z - Kirill Eremenko)

During training, Encoder is fed with Questions and decoder with Answers. Weights in gates, hidden states get settled. During testing for each sequence of input, encoder results in to a combo vector. Decoder takes this and starts spitting out words one by one, probabilistically.

YHK

# Transformers

- The Transformer  is a model that uses attention to boost the speed with which seq2seq with attention models can be trained. The biggest benefit, however, comes from how The Transformer lends itself to parallelization.
- In its heart it contains an encoding component, a decoding component, and connections between them.




YHK

# Transformers

- Offer a better structure to train a language model, which gave raise to the large language models (LLMs) like GPT and Bart. Its characteristics are:
  - Positional encoding: each word is labeled with the number of its position in a sentence.
  - Self-attention: each word is examined in the context of the whole sentence to generate a representation of the word. This helps the model to understand the linguistic meaning and nuances of a word. As the scale of a language model grows, the model builds mastery of our human language, and it does not only know how to perform basic text-based tasks but also gives a structured and logical answer to any user prompt.

(Ref: Techy Stuff 1: Notes on Transformers, LLMs, and OpenAI - Bill)

YHK

# GPT

# Generative Pretrained Transformer (GPT)

2018's GPT was a big success in pretraining a decoder!

- Transformer decoder with 12 layers.
- 768-dimensional hidden states, 3072-dimensional feed-forward hidden layers.
- Trained on BooksCorpus: over 7000 unique books.
- Contains long spans of contiguous text, for learning long-distance dependencies.
- The acronym "GPT" never showed up in the original paper; it could stand for
- "Generative PreTraining" or "Generative Pretrained Transformer"

(Ref: John Hewitt, Radford et al., 2018)

YHK

# GPT-3, in-context learning, very large models

- Very large language models seem to perform some kind of learning without gradient steps simply from examples you provide within their contexts.
- GPT-3 is the canonical example of this. The largest T5 model had 11 billion parameters.
- GPT-3 has 175 billion parameters.

(Ref: John Hewitt)

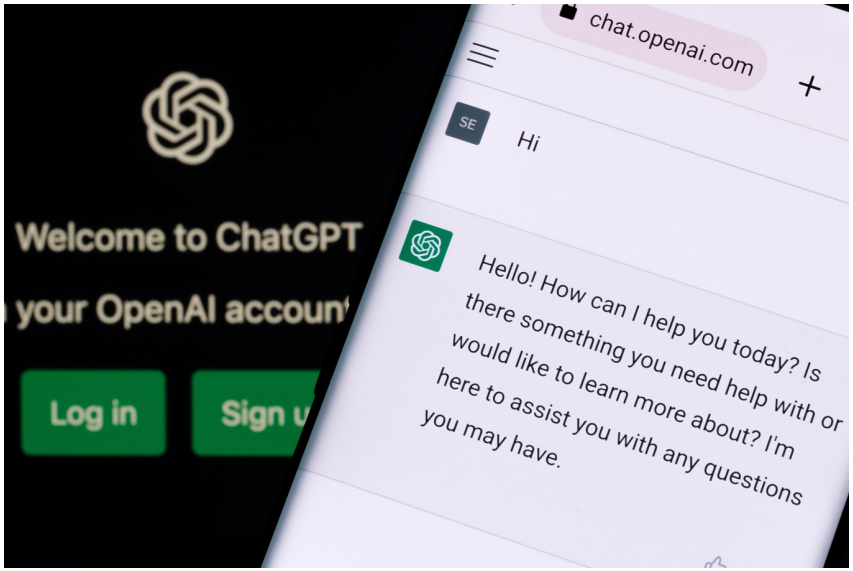
YHK

# Overview of ChatGPT

YHK

# What is ChatGPT?

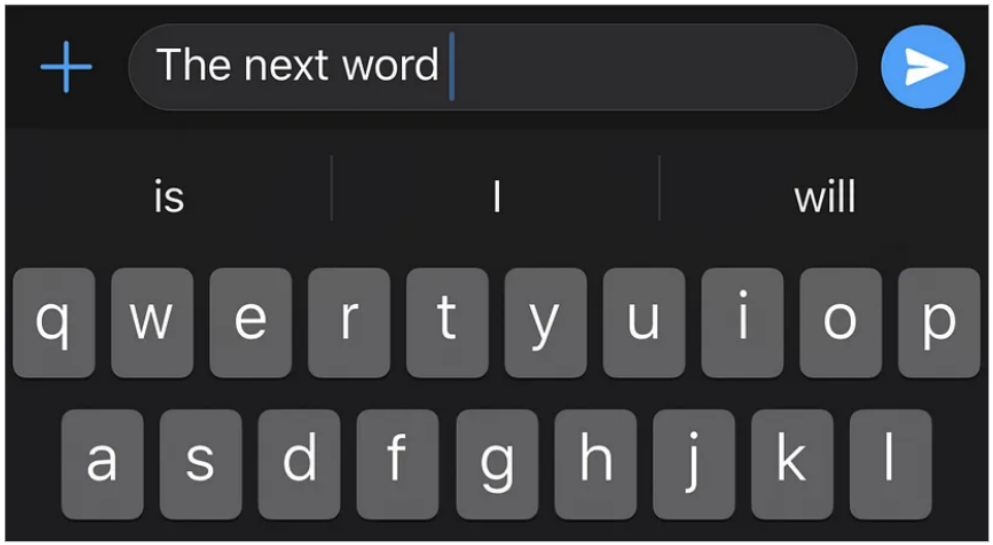
- A Chatbot
- A language model, that takes 'prompt' from user and generates a response.
- Built by OpenAI and released in Nov 2022
- Got 1m users in 5 days (Insta 2.5 months, Spotify 5m, Facebook 10m, Netflix 3.5yrs)
- Link: https://chat.openai.com/chat
- Details: https://openai.com/blog/chatgpt/




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

# What is a Language Models?

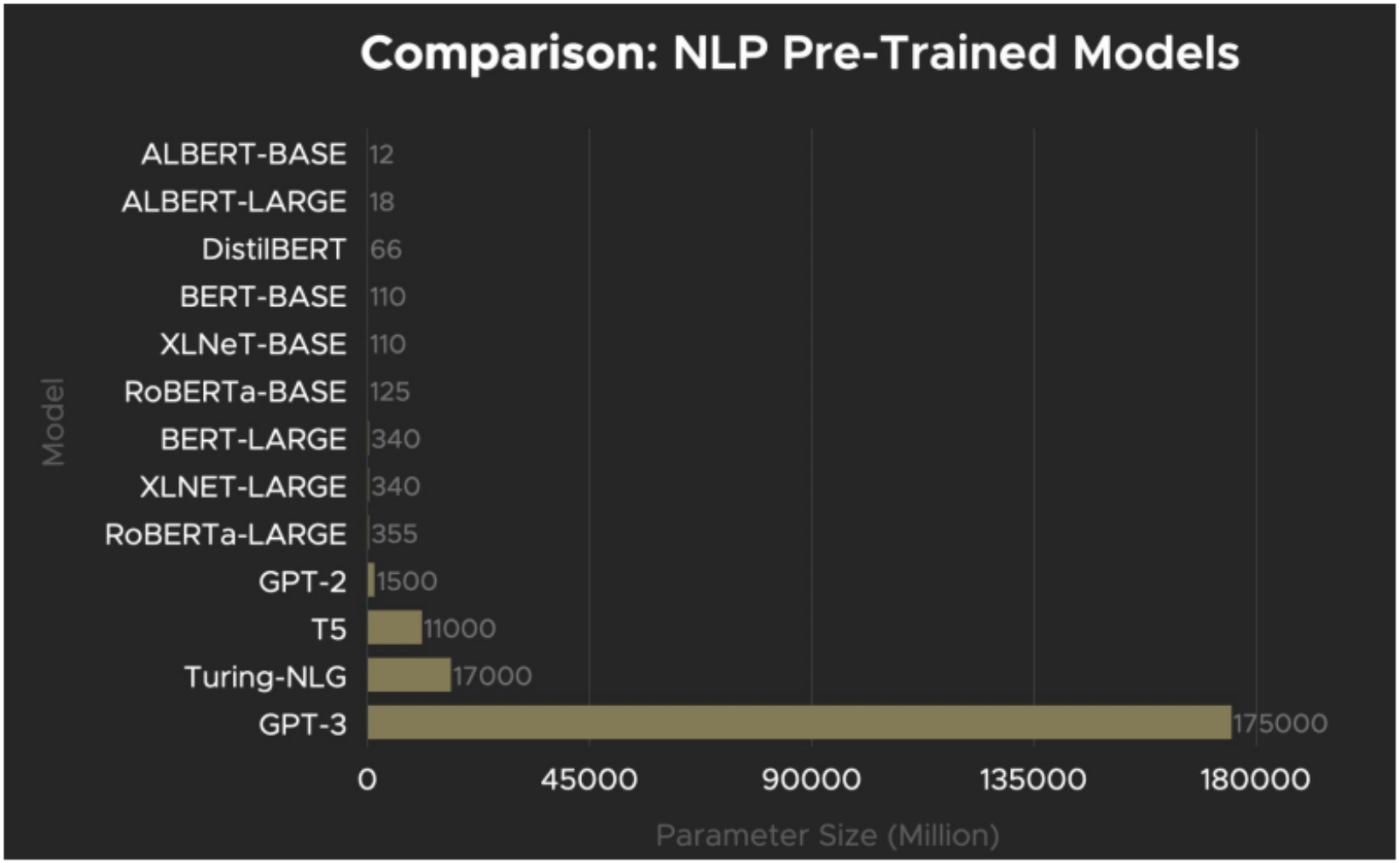
- While typing SMS, have you seen it suggests next word?
- While typing email, have you seen next few words are suggested?
- How does it suggest? (suggestions are not random, right?)
- In the past, for "Lets go for a . . . ', if you have typed 'coffee' 15 times, 'movie' say 4 times, then it learns that. 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

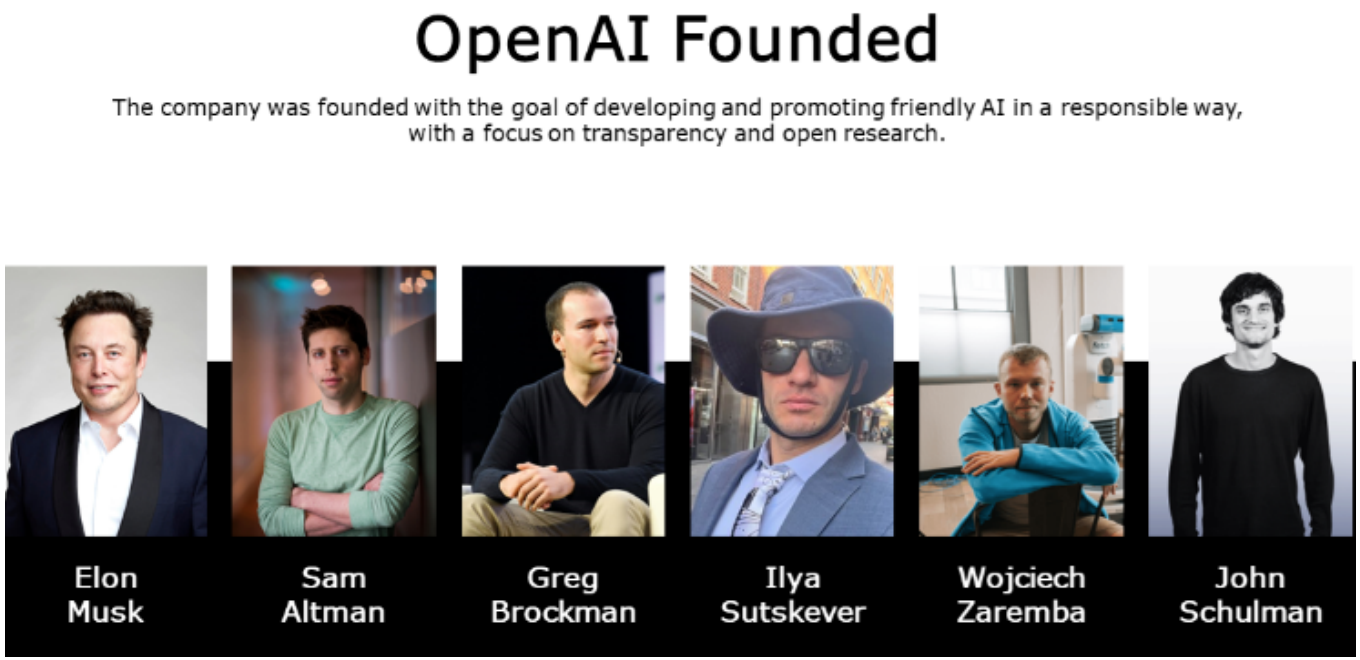



(Ref: Deus.ai https://www.deus.ai/post/gpt-3-what-is-all-the-excitement-about)

YHK

# 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




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

YHK

# 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

YHK

# 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)

YHK

# GPT3 vs ChatGPT

| | GPT-3 | ChatGPT |
|---|---|---|
| **Training Parameters** | 175 billion | 175 BillionRLHF TechniqueHuman conversational text |
| **Designed For** | NLP Tasks | Chatbot apps |
| **Content Retention Ability** | No | Yes |
| **Use Cases** | 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)

YHK

# Conclusions

# Advantages

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

YHK

# 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

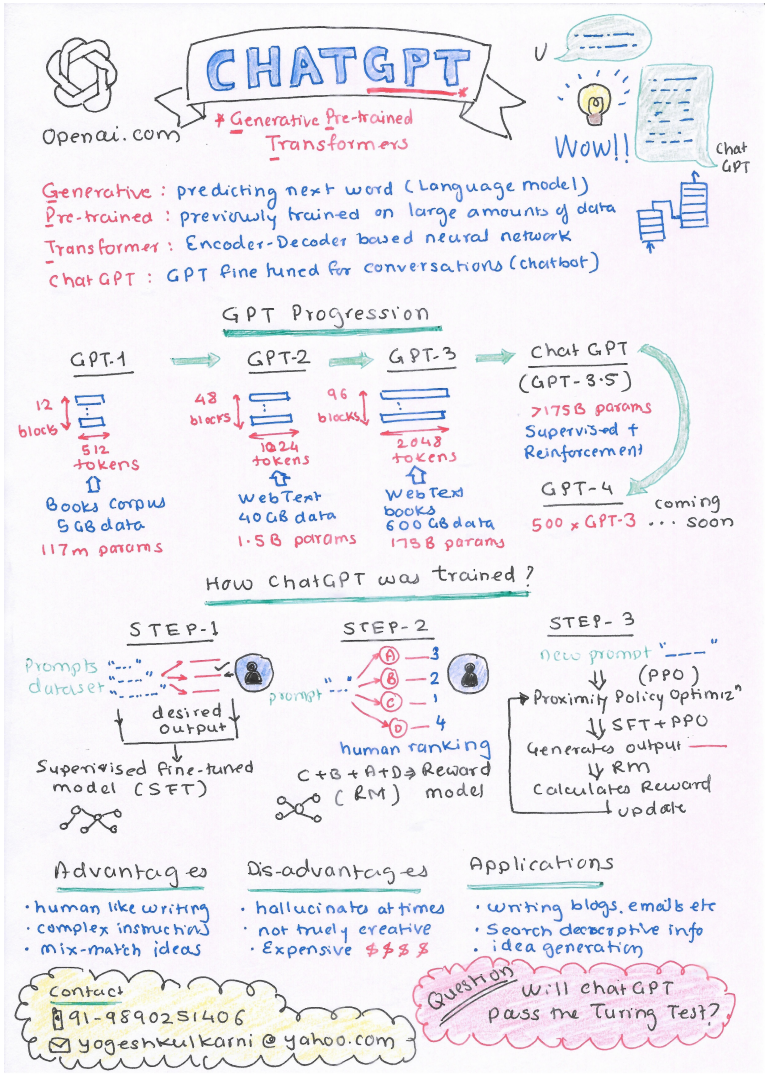
YHK

# 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

YHK

# My Sketchnote




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

# The After-shocks

YHK

# 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)

YHK

# Will ChatGPT Kill Jobs?

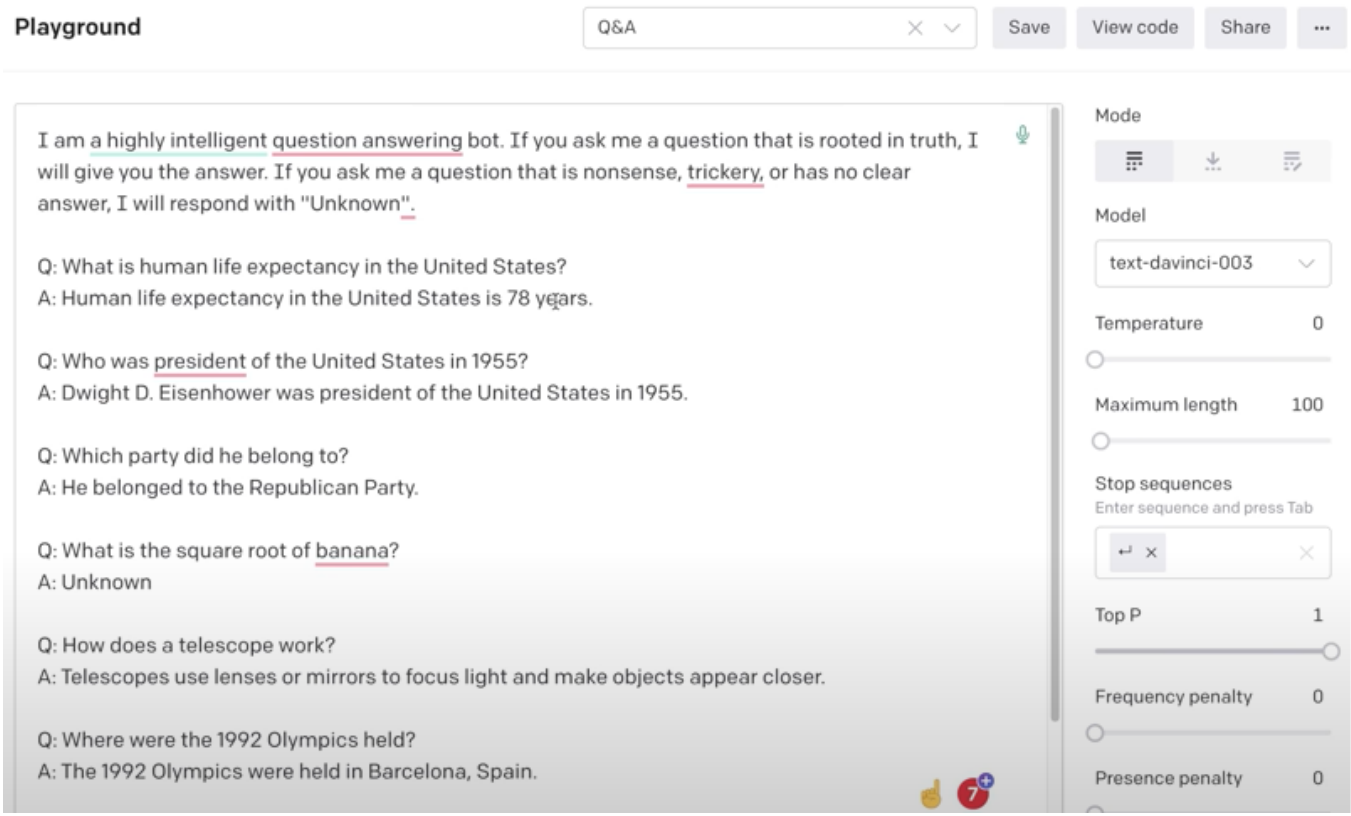

Repetitive, boring and standard, language based jobs, for sure. Need to be more creative, experiential to stand against ChatGPT.

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

YHK

# 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)




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

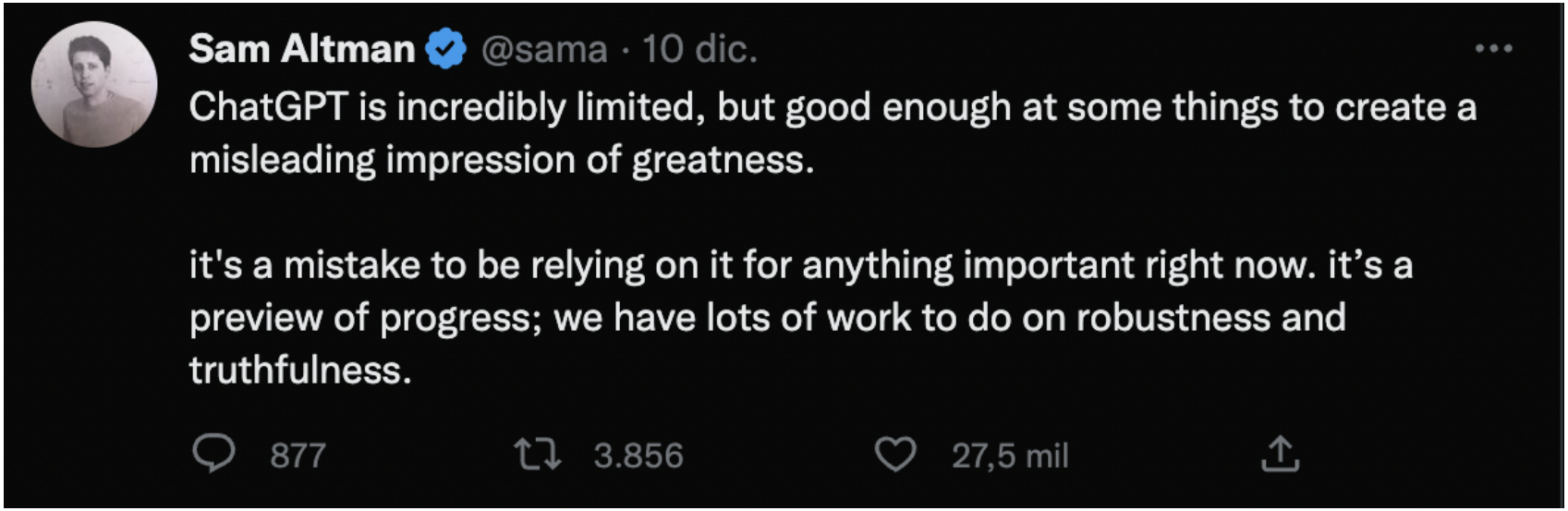
YHK

# The Hype


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

YHK

# Finally, from horses mouth!!




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

YHK

# References


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


YHK

# 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

(Generated by Hugging Face QR-code-AI-art-generator, with prompt as "Follow me")

YHK