RESEARCH

TECHNOLOGY BEHIND CHAT-GPT

BY

VIDHI VERMA

RASHMI BAGHEL

SANDEEP

2023

Technology behind ChatGPT

Vidhi Verma | Rashmi Baghel | Sandeep

Vidhiverma0910@gmail.com | baghelrashmi1@gmail.com | sk1952978@gmail.com

8368765396 | 9990988334 | 7982758858

Abstract

This research paper discusses the technology behind ChatGPT, and how capable this conversational AI art is. The objective of this study is to analyze all the technologies which are tightly interwoven and responsible for creating human-like responses at the back. To analyze the training methods including initializing the model, fine-tuning a model, and the application of ChatGPT in the field of NLP. In our research paper, we have given a brief about the main algorithm of ChatGPT 3.5 architecture which is NLPO and RLHF, and main technology and models like Deep learning, Machine learning (supervised learning, unsupervised learning, and Reinforcement learning), Neural networks, Cloud computing, etc. About ChatGPT, GPT referring to the generative pre-trained transformer was first launched back in 2018 with 117 parameters, GPT-2 in 2019 with 1.5 billion parameters, and GPT-3 in 2020 largest model ever built with 175 billion parameters our findings state that ChatGPT exhibits impressive capabilities in understanding and generating human-like responses. Having proficiency in various conversational tasks, such as answering questions, engaging in dialogue, and providing useful information. But being an AI model, limitations were observed in handling ambiguous as well as simple queries many times and maintaining consistent context over long conversations. In conclusion, our study through this research paper provides a comprehensive analysis of the technology behind ChatGPT, showcasing its strengths and limitations in conversational AI.

Keywords

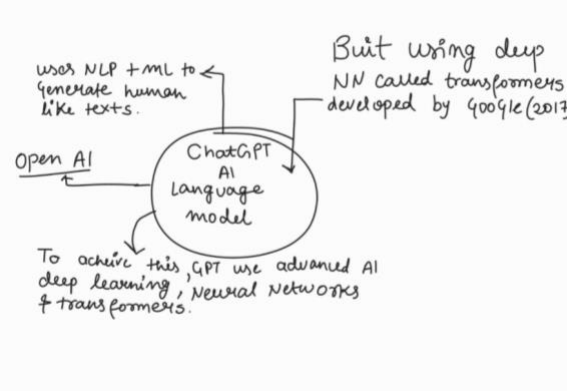
RLHF (Reinforcement Learning from Human Feedback), NLPO (natural language policy optimization), MLM (Masked Language Modeling), PPO (Proximal Policy Optimization), transformers, language modeling, large-scale data processing, neural networks, deep learning, machine learning, word embedding, attention mechanism, gradient descent optimization algorithm, graph neural networks, convolutional neural networks, recurrent neural networks, Autoencoders.

Introduction

ChatGPT is a sophisticated AI developed by OpenAI, which uses several technologies and proves to be a monster chatbot that can do almost everything required to improve efficiency in healthcare, finances, marketing, education, innovation, and many other tasks. ChatGPT generates codes, and can do editing, Q&A, and composition, for which at its core, OpenAI built some language models under the GPT series, including GPT, GPT-2, GPT-3,

GPT-NEO, GPT-NEO2, and Codex. Some transformer-based languages models like GPT-2, GPT-3, BERT (designed by Google and fine-tuned by OpenAI), Transformer, CTRL, and GShard are used to produce coherent and fluent generation. The Transformers family uses a self-attention mechanism to produce input data and includes both encoder and decoder architecture.

The technical aspects of ChatGPT, neural networks, GANs, deep learning, and machine learning interwoven with RL and NLPO involve training, pre-training, fine-tuning, evaluation, and deployment of a model. Cloud computing is another crucial technology used in ChatGPT, as it not only provides infrastructure but also acts as a bridge between other cloud stations. We have also given some insights about the future of ChatGPT and the head-to-head AI race, a technology-driven competition between ChatGPT and Google's AI Bard.



Algorithms

For training conversational agents, NLPO and RLHF algorithms are majorly used. To update the model continuously based on the human feedback received is done by RLHF. On the other hand, NLPO generates human-like responses using the datasets and patterns/structures learned from the feedback. NLPO does not require any human feedback because it uses RL for better optimization. RLHF and NLPO both the algorithms require training processes for better results and performance. This training is done on small datasets and a large corpus of text data to understand and make the model able to work on datasets. These training methods are fine-tuning and pre-training respectively.

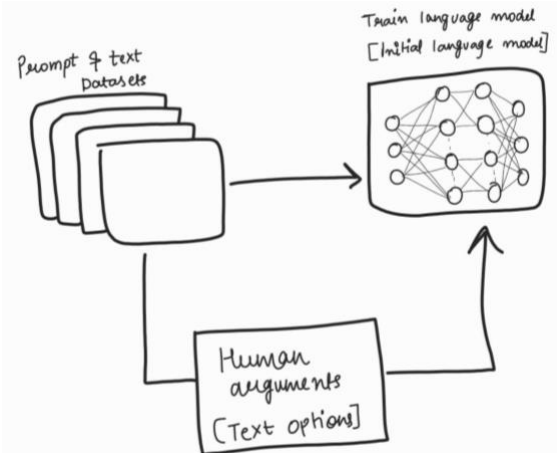
In brief, we can say that RLHF requires human feedback to update and increase the performance and NLPO does not require any human feedback as it optimizes reinforcement learning for that.

Algorithm 1 [RLHF]

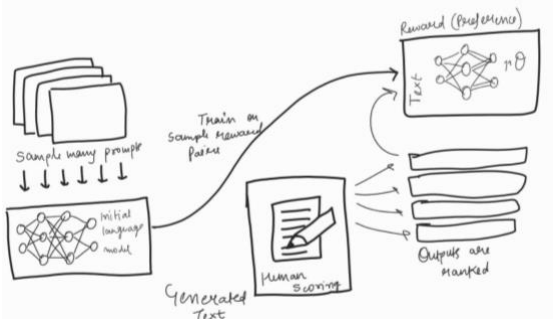
The concept of reinforcement learning from human feedback (RLHF), involves a multi-step training process and different stages of deployment.

The training process is broken down into three core steps:

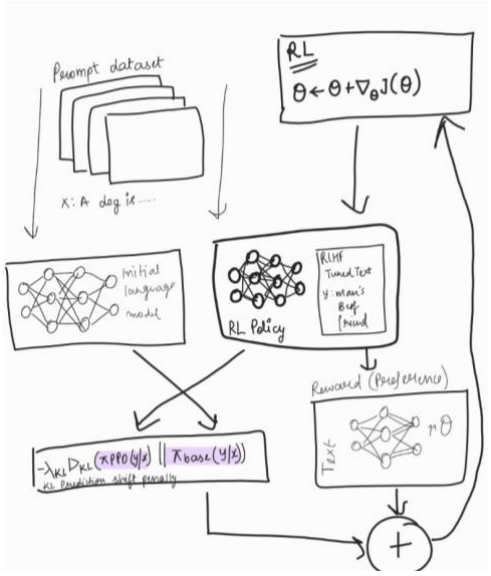
Pretraining a language model (LM), gathering data and training a reward model, and fine-tuning the LM with reinforcement learning.[2]



Reward models are generated using human preferences, and how reinforcement learning is used to optimize the initial language model concerning the reward model. [2]



Proximal Policy Optimization (PPO) algorithm for fine-tuning. The design space of options in RLHF training is yet to be thoroughly explored, and the article highlights some of the challenges in this area. [2]



The Reinforcement Learning from Human Feedback (RLHF) algorithm involves training an AI model using a combination of supervised learning and reinforcement learning. Here are the steps involved:

Initialize the model: The algorithm starts by initializing a model using a supervised learning approach, using a dataset of human-generated dialogue examples.[2]

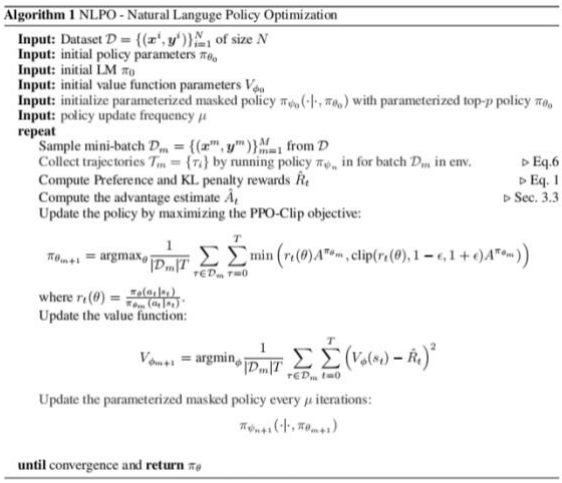
Fine-tune the model: The model is then fine-tuned using RLHF. Human trainers play the role of both user and AI assistant, and the model generates suggestions to help the trainers craft their responses. The resulting dialogue dataset is combined with an existing dataset and transformed into a dialogue format.[2]

Collect comparison data: To create a reward model for reinforcement learning, the researchers collect comparison data by having human trainers rank different responses generated by the model. They randomly select a model-written message and sample several alternative completions, which human trainers then rank.[2]

Train using reinforcement learning: Using the reward models, the model is fine-tuned using Proximal Policy Optimization (PPO), which is a reinforcement learning algorithm. Several iterations of this process are performed until convergence.[2]

Overall, the RLHF algorithm is designed to allow an AI model to learn from human feedback more efficiently and effectively, by combining supervised and reinforcement learning approaches.

Algorithm 2 [NLPO]



Algorithm 1 NLPO - Natural Languge Policy Optimization

Input: Dataset $\mathcal{D} = \{(x^i, y^i)\}_{i=1}^{N}$ of size N
Input: initial policy parameters π_{θ_0}
Input: initial LM π_0
Input: initial value function parameters V_{ϕ_0}
Input: initialize parameterized masked policy $\pi_{\psi_0}(\cdot|\cdot, \pi_{\theta_0})$ with parameterized top-p policy π_{θ_0}
Input: policy update frequency μ
repeat
Sample mini-batch $\mathcal{D}_m = \{(x^m, y^m)\}_{m=1}^{M}$ from $\mathcal{D}$
Collect trajectories $T_m = \{\tau_i\}$ by running policy π_{ψ_n} in for batch $\mathcal{D}_m$ in env. ▷ Eq.6
Compute Preference and KL penalty rewards $\hat{R}_t$ ▷ Eq. 1
Compute the advantage estimate $\hat{A}_t$ ▷ Sec. 3.3
Update the policy by maximizing the PPO-Clip objective:

$$\pi_{\theta_{m+1}} = \text{argmax}_\theta \frac{1}{|\mathcal{D}_m|T} \sum_{\tau \in \mathcal{D}_m} \sum_{\tau=0}^{T} \min\left(r_t(\theta) A^{\pi_{\theta_m}}, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A^{\pi_{\theta_m}}\right)$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_m}(a_t|s_t)}$.
Update the value function:

$$V_{\phi_{m+1}} = \text{argmin}_\phi \frac{1}{|\mathcal{D}_m|T} \sum_{\tau \in \mathcal{D}_m} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t\right)^2$$

Update the parameterized masked policy every μ iterations:

$$\pi_{\psi_{n+1}}(\cdot|\cdot, \pi_{\theta_{m+1}})$$

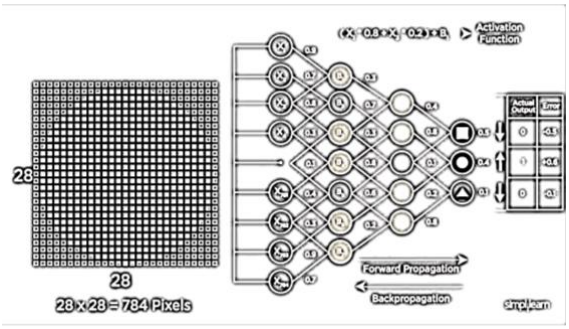
until convergence and **return** π_θ

This is an algorithm for natural language policy optimization. It takes a dataset of input-output pairs, an initial policy, an initial language model, and an initial value function as inputs. It initializes a masked policy and updates it by collecting trajectories, computing rewards, estimating advantages, and maximizing the PPO-Clip objective. The value function is also updated, and the parameterized masked policy is updated every μ iteration. This process is repeated until convergence, and the final policy is returned.

Neural networks

Neural networks in ChatGPT are inspired by the biological functioning and structure of the human brain. The more data we have, the more efficient performance will neural network show. It consists of billions of nodes or neurons connected performing forward and backpropagation to predict more accurate outputs in this multi-layered network. In terms of mechanism, each artificial neuron receives input signals from the other neuron or external environment [like having stimuli] for a reaction, and then this network performs mathematical functions on the input to extract an output signal. A further output of this neuron is fed to another neuron as an input. By adjusting the weights of connections between neural networks, it learns to recognize patterns to make predictions. here weights are the

numerical values assigned between the neuron layers. The source of this massive data for artificial neural networks are books, websites, articles, news, and publicly available text sources.

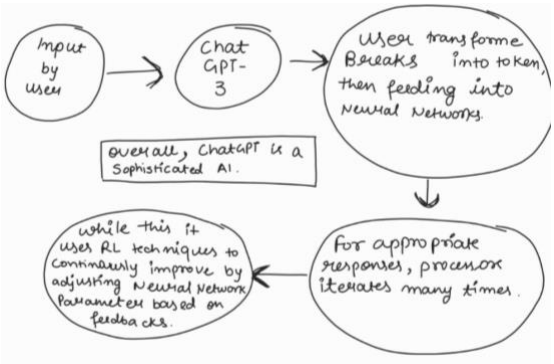


[1]

Types of neural networks used in ChatGPT are CNN [convolution neural network], GNN [graph neural network], RNN [recurrent neural networks], DNN [deep neural networks], ANN [artificial neural networks], LSTM [long-short term memory]. Apart from this, GPT uses neural networks depending on different specific tasks such as GANs [generative adverse network which is used for generating new data], and autoencoders [used for unsupervised learning and dimensionality reduction.].

Deep – learning

"Deep" represents here neural networks which consist of many layers of neurons enabling it to understand and learn complex representations of input data. It primarily feeds massive data into a neural network & automatically learns specific features. Diving into the mechanism, while training, the weights and biases (parameters) are adjusted by optimizing algorithms such as stochastic gradient descent. The end goal of deep learning is the reduction of loss function which measures the difference between the prediction (predicted output) and reality (actual output). By doing this, the network learns to make more accurate predictions on unseen data. Deep learning involves majorly four parts:



1. Data preparation, where it collects data & then splits it into training & testing sets.
2. Neural network architecture, which involves the arrangement of nodes. The most common type of deep learning is a "feedforwarding neural network".
3. Training the network, involving forward and backpropagation where weights are needed to be adjusted.
4. Testing and validation, after which the model is deployed on unseen data.

Machine – learning

Machine learning {Consists of algorithms to learn patterns in data and make predictions or decisions}. Machine learning involves fewer complex algorithms than those deep learning. Like deep learning, it also involves training models, but it doesn't require as much computing power to train as deep learning. In terms of mechanism, it differs from deep learning in complexity, depth of models, amount of data, and computing power. Machine learning has three types:

1. Supervised learning (SL): which makes accurate predictions, and involves training models with labeled datasets. The process of supervised learning includes **data collection**, where each point has an input and a corresponding output, **and model selection**, which is more about choosing an appropriate machine learning model for the specific problem, for example, a regression model, a decision tree, a support vector machine (SVM), or other algorithms. **Training** involves

training the model on labeled datasets. **Evaluations** involve evaluating the performance of our pre-trained models on separate or unseen data. **Deployment** happens when the model is trained and evaluated, and it's ready to make predictions, for example, classifying emails as spam or not spam based on their content.

2. Unsupervised learning: Unlike supervised learning, which predicts accurate and precise outputs/data, unsupervised learning involves training models with unstructured, unlabeled raw data. Hence the computer sorts the data without having any instructions about what to do and what not to do, unlike SL. It finds out patterns and similarities on its own and tries to make sense of the dataset as a whole. For example, clustering, dimensionality reduction, anomaly detection, and generating modeling.

3. Reinforcement learning: As discussed above in the algorithms section, RL trains a model with state, action, and reward signals. Synthesizing, RL includes pre-training a model, reward models, and proximal policy optimization (PPO) algorithm for fine-tuning. Although RL is not a primary technology, it can be considered a crucial component. Reinforcement learning works on various criteria such as relevance, coherence, and novelty.

Cloud – computing

Cloud computing and its technologies have improved AI a lot. We can see dynamic forces of AI that are data and datasets, processing capacity including GPUs, and other skills. Cloud computing is enhancing AI with its features. Infrastructure as a service [component of Cloud computing] made the AI capable of providing an infrastructure environment and GPU [graphics processing unit] to the user. Platform as a service provides data science services to AI. Like Jupyter notebooks. Software as a service provides software services to the AI like payment applications for better results.

Whatever we discussed above are the key technologies behind ChatGPT. Overall, ChatGPT has a huge technical background that involves Transformers [a type of neural network that is proven successful for the implementation of NLPO.], Large–scale data processing [processing and analyzing big data sets.], programming like Python, c++, and Java, Disturbed Computing Frameworks [Hadoop, Apache spark], Unsupervised learning [the learning that allows the AI to find patterns and structures without any explicit targets.], Transfer learning [the process of transforming pre-trained models and helping them in learning new tasks and domains.], Ensembling techniques [the technique of combining multiple models into one to improve performance.], Word embeddings [vector representation of words.], Attention mechanisms [to weigh the importance of different inputs and tokens.], Text processing techniques [to make the data more clean and suitable for machine learning models.], Gradient descent optimization [to teach the machine about iterative parameters.], Regularization techniques [the process of adding constraints to a model,], Evaluation metrics [used to measure the performance of the machine.], Names identity reorganization [used to identify named entities in text. Like the name of a person or name of an organization.], Part-of-speech tagging [labeling each part of a sentence: nouns, verbs, adjectives, etc.], Semantic role labeling. ChatGPT is knowingly using around 50 technologies and algorithms behind it. All these technologies together make ChatGPT able to learn data, understand human queries, and provide human-like responses to them.

What next in technology for ChatGPT?

The techniques, such as RLHF and NLPO are in use by many AI research labs and models

currently due to their heavy benefits and better performance in this field. But, this agathism and techniques still hold limitations. The biggest limitation is that the models using these techniques can provide output that is harmful or inaccurate as they are learning from the big data. This can lead to the inherent human problem challenge in the future. The collection of data from workers outside the training loop is also a limitation of these techniques. The quality of RLHF's performance depends on the quality of annotation provided by the human which leads to the need for training for the user also. Training part-time staff and hiring them can cause high costs. Despite what has just been said, the training model of RLHF is not as expensive. Hence, it's not clear whether the benefits of these technologies cover the limitations. But, further research and experiments will surely update and upgrade all these technologies and will be a better version by integrating human feedback into natural language models. The future remains suspenseful, but the spoiler we have is that AI and its technologies will keep evolving and become better with every research and update.

Limitations of ChatGPT

As mentioned earlier Despite its impressive capabilities, ChatGPT does have certain limitations observed in its functionality. Due to the absence of real-time interaction and the reliance on pre-trained models, ChatGPT may encounter challenges in disambiguating queries with multiple potential interpretations. Consequently, the responses generated by ChatGPT in such scenarios might be incomplete or lack contextual appropriateness, which could lead to potential misunderstandings.

Furthermore, maintaining consistent context over longer conversations poses a challenge for ChatGPT. The model operates within a fixed window and lacks continuous memory of previous interactions. As a result, maintaining coherence and understanding in extended conversations is proved to be difficult for ChatGPT.

Another important limitation to consider is the presence of biases in ChatGPT's responses due to biases in its training data. Since the model learns from large-scale datasets that inherently reflect societal biases and preferences, it can generate biased responses or exhibit favoritism towards certain groups. These biases have ethical implications and can impact the overall trustworthiness of ChatGPT's outputs.

The model is primarily trained on general internet text, which may limit its ability to provide accurate or relevant responses in specialized or uncommon domains.

Google's BARD vs ChatGPT

In this AI race and the rise of ChatGPT in the field of AI, Google is not going to sit and watch. Google's BARD is the upcoming AI rival for ChatGPT. Bard works on LaMDA [language model for dialogue application], the transformer base neural language that is trained on online chat data via pre-training models. Bard uses and follows the bottom-up approach and ensures better speed and the highest level of correctness in data by running the data from multiple matrixes. Bard currently holds 1.37 billion parameters that are comparatively very less than ChatGPT's parameters [175 billion parameters]. Bard is capable of generating audio responses in different voices. whereas, ChatGPT ensures its more flexible and unpredictable responses. Google won't let GPT collect all the fame and name in the NLP market alone. hence, Google's Bard is designed to beat ChatGPT and provide better performance to its users. As of now, google bard is free of cost and will be available publicly soon. but GPT is leading in the NLP market now. yes, both the AIs are fighting to become the best but it's also true that both the AIs do have some limitations.

more upgrades and developments will keep making them better. but an unavoidable and unwanted war has started between these two AIs. just like Apple and iPhone, iOS and Android, Ferrari, and Ford, and there is no lack of examples.

Conclusion

In conclusion, after lots of research and development in the field of NLP and AI, the technology behind ChatGPT has come up so far. ChatGPT is a result of a complex combination of techniques, algorithms, programming, data sets, DL and ML, training, and big data to make it able to understand and generate human-like responses. After lots of experiments and training processes, the developers of ChatGPT have made it able to understand and generate real-time query responses and made it work like personalized assistance with lots of data information. The research and development so far made it able to help and provide tools and data to people all around the world. Yet, there is Soo much more to learn and develop in this field of AI. The research, experiments, and developments will keep upgrading these technologies and keep presenting a better model in front of the users.

References

[1] Neural Network In 5 Minutes | What Is A Neural Network? | How Neural Networks Work | Simplilearn | Youtube
https://www.youtube.com/watch?v=bfmFfD2RIcg

[2] Illustrating Reinforcement Learning from Human Feedback (RLHF)
Published December 9, 2022
https://huggingface.co/blog/rlhf

[3] DEFINITION natural language processing (NLP) By Ben Lutkevich, Technical Features Writer
https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP

[4] Chatting about ChatGPT: How may AI and GPT impact academia and libraries?
Article in Library Hi Tech News · January 2023
https://www.researchgate.net/publication/367161545

[5] The AI Race is on! Google's Bard and OpenAI's ChatGPT Head to Head: An Opinion Article 6 Pages Posted: 8 Feb 2023 by Md. Saidur Rahaman Metropolitan University, Sylhet, Bangladesh
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4351785