

[Home](#)[My Network](#)[Jobs](#)[Messaging](#)[Notifications](#)[Me](#)[Tools](#)[Learn](#)

Yogesh Haribhau Kulkarni

Principal Architect, CTO Office, Icertis | PhD in Geometric Modeling | Adjunct Professor (Data Science, CoEP) | Google Developer Expert (Machine Learning) | Top Writer 3x (Medium) | ChatGPT-LangChain

[View full profile](#)

Yogesh Haribhau Kulkarni • You

Principal Architect, CTO Office, Icertis | PhD in Geometric Modeling | Adjunct Profes...
2d • Edited •

Custom Training for Language Models

Although pretrained models offer out-of-the-box functionality for many NLP (Natural Language Processing) tasks, but they go only that far. Soon, a need arises to train on your own data (custom corpus). The ways in which custom training is done, have changed over time, you can see this progression, below:

[1] Own data only: Frequency-based models such as bag-of-words and tf-idf (term frequency inverse document frequency) are typically trained from scratch, fully on the custom corpus. No pretrained models as such. Use them as vectorizers in any NLP tasks such as classification, via machine learning or deep learning neural networks.

[2] Re-training: Model like word2vec comes in pretrained flavor from Google. You can use it as a vectorizer in any NLP tasks, as mentioned above. But, Python library Gensim, allows you to build your own word2vec (and doc2vec) model from scratch also, fully based on your own corpus. This gives you your own word-vectors to be used in any NLP tasks, as mentioned above.

[3] Fine-tuning: Models like BERT (Bidirectional Encoder Representation of Transformers) come in pretrained flavor to be used as a vectorizer and also as base neural network. Based on downstream NLP tasks, such as classification, one needs to add, last layer to that base neural network and thus you get your own classifier which can be trained on the custom corpus.

[4] Prompt-tuning: Models like *GPT* (Generative Pretrained Transformers), which are LLMs (Large Language Models), are so powerful, that there is no need to tamper the neural network for fine tuning with your own data. You just need to give a few samples (shots) within the prompt itself and it starts answering based on the samples. Of course, one can do all the above methods as well, but this new way of fine-tuning via prompts has made LLMs accessible to anyone-and-everyone.

Riding on the wave of this approach of custom training for language models, new frameworks/platforms are emerging that allow you to prompt/fine-tune LLMs easily.

LangChain is one of them. Google will soon have a similar offering in Cloud integrated with Vertex AI. You have already started seeing such custom GPTs, like GitaGPT!!

Such solutions are going to give birth to a new category, called 'SaaS-GPT' (Software-as-a-Service based on GPT).

Watch this space for more updates!!

[#gpt](#) [#chatgpt](#) [#ai](#) [#nlp](#) [#saas](#) [#saasgpt](#) [Google](#) [LangChain](#) [#langchain](#) [#openai](#) [#vertexai](#) [#googlecloud](#) [#gensim](#) [OpenAI](#) [#deeplearning](#) [#neuralnetworks](#) [#machinelearning](#) [#data](#) [#software](#) [#python](#) [#cloud](#) [#google](#) [#training](#) [#language](#)



Akash Suresh Mavle and 41 others

2 comments

Reactions



Like



Comment



Repost



Send

4,265 impressions

[View analytics](#)



Add a comment...



Most relevant ▾



Yogesh Haribhau Kulkarni • Author

2d ...

Principal Architect, CTO Office, Icertis | PhD in Geometric Modeling | Adjunct Professor (Data Science, CoEP) | Google Developer Expert (Machine Learning) | Top Writer 3x (Medium) | ChatGPT-LangChain

Corresponding [Medium](#) story, with ref links at <https://medium.com/p/84ba80265d0f>

Like | Reply



Shreyas Dixit • 1st

2d ...

Student Researcher | Deep Learning | Joint Technical Head @SEC VIIT | ML Developer @MLSC VIIT

I have also built a similar GPT model series called ScriptGPT. It allows users to generate video scripts from titles. It is currently trained on only 250 videos as a dataset for this is yet to be built and also built on local GPU.

ScriptGPT: https://huggingface.co/SRDdev/Script_GPT

...see more

Like · 5 | Reply

Ad ...

Keep up with interesting, relevant updates



Yogesh Haribhau, grow your business with news and insights from Honeywell Forge

[Follow](#)