GPT Mastery: A Comprehensive Guide To Crafting Your Own ChatGPT

Week 6

1.) State of GPT | BRK216HFS

In this video, we will explore the training pipeline of GPT assistants, such as ChatGPT. We'll cover the process from tokenization to pretraining, supervised finetuning, and Reinforcement Learning from Human Feedback (RLHF). Additionally, we'll delve into practical techniques and mental models for effectively using these models, including prompting strategies, finetuning, the rapidly growing ecosystem of tools, and their future extensions.

Youtube: ▶ State of GPT | BRK216HFS

2.) Let's build the GPT Tokenizer

In this video, we will explore the critical role of the Tokenizer in Large Language Models (LLMs). The Tokenizer converts between strings and tokens (text chunks), acting as an essential component of the LLM pipeline. This process involves its own training sets and algorithms, like Byte Pair Encoding. After training, the Tokenizer performs two key functions: encoding strings into tokens and decoding tokens back into strings. We will build the Tokenizer used in the GPT series from OpenAI from scratch, uncovering how many peculiar behaviors and issues in LLMs originate from tokenization. We'll discuss these problems, why tokenization is to blame, and consider why it might be beneficial to eliminate this stage altogether.

Youtube: ▶ **Let's build the GPT Tokenizer**
Github:https://github.com/karpathy/nn-zero-to-hero