Term	Definition
Direct preference optimization (DPO)	An optimization technique that leverages pairwise comparisons (preferences) rather than explicit rewards or scores to train models, especially useful in scenarios where assigning precise numerical scores is challenging.
Distribution (in ML)	A function that shows all the possible values of a data set and how often they occur. In the context of language models, it refers to the probability distribution of different possible responses given an input query.
Fine-tuning	The process of adapting a pre-trained model to a specific task or dataset by continuing the training process on new data. Fine-tuning allows models to achieve better performance on the specific task by leveraging existing knowledge from pre-training.
Hugging Face	A platform that provides tools, libraries, and resources for building, training, and deploying machine learning models, especially those based on transformers like GPT-2.
IMDB dataset	A dataset containing 50,000 movie reviews used for sentiment analysis, commonly used to train models to classify reviews as positive or negative.
Inference	The process of using a trained machine learning model to make predictions or generate outputs based on new input data. In the context of language models, inference refers to generating text or making predictions using the trained model.
Kullback-Leibler (KL) divergence	A measure of how one probability distribution diverges from a second, reference probability distribution. It is often used to ensure that the new policy remains close to the old policy during training in reinforcement learning.
Language model	A model that predicts the probability of a sequence of words. It is used in various applications, including generating responses in conversational AI based on an input query.
LengthSampler	A method used to vary text lengths for data processing in machine learning models, enhancing robustness and simulating realistic training conditions by managing input text lengths.
Log-derivative trick	A mathematical technique used to calculate the gradient of a function when the function itself is given in an expectation form, often used in reinforcement learning to optimize policies.
Low-rank adaptation (LoRA)	A technique for parameter-efficient fine-tuning, particularly in transformer models, that adds trainable low-rank matrices to each layer of a pre-trained model to reduce the computational cost of training and to make fine-tuning more memory-efficient.
Loss function	A function that measures the difference between the predicted outcomes of a model and the actual target values. It is used to guide the optimization of the model by minimizing this difference during training.
Max and min tokens	Parameters that set the maximum or minimum number of tokens generated in a sequence, used to control the length of outputs in language models.