Term	Definition
Objective function	A mathematical function that represents the goal of an optimization problem, often used in machine learning to guide models toward better performance by minimizing or maximizing this function.
Omega (ω) function	A notation used to describe the detailed policy distribution in reinforcement learning, representing how the probabilities of a response are computed based on the previous tokens in the input sequence.
Optimization	The process of adjusting the parameters of a model to improve its performance, typically by minimizing a loss function or maximizing a reward. In DPO, optimization involves maximizing the log-likelihood of the DPO loss.
Partition function	A mathematical function that sums over all possible states or outcomes of a system, often used in statistical mechanics and in the normalization of probability distributions in machine learning, especially in reinforcement learning where the number of possible outcomes can grow exponentially.
Pi (π) policy	In reinforcement learning, the policy that defines the probability distribution over actions given a state, often denoted as π. It is the model that is optimized to achieve the best performance in decision-making tasks.
Pipe outputs list	A list that stores the outputs of a pipeline in Hugging Face, particularly in the context of sentiment analysis, where it contains the sentiment scores for generated responses.
Policy gradient	A method in reinforcement learning that optimizes the policy directly by maximizing the expected reward.
PPO config class	A class used to specify the model and learning rate for proximal policy optimization (PPO) training, defining essential configurations for training models.
PPO trainer	A specialized trainer in reinforcement learning that processes query samples, optimizes chatbot policies, and handles complex tasks to ensure high-quality responses.
Proximal policy optimization (PPO)	A reinforcement learning algorithm that optimizes the policy of an agent by ensuring that updates are not too drastic, thus stabilizing the training process.
Reinforcement learning from human feedback (RLHF)	A reinforcement learning technique where human feedback is used to guide the learning process of models, particularly useful in optimizing large language models for tasks like chatbots and recommendation systems.
Repetition penalty	A parameter that penalizes repeated sequences of tokens during text generation, encouraging more diverse outputs and reducing the likelihood of generating repetitive content.
Reference model	A pre-trained model used as a baseline or comparison point in further training or optimization, particularly in reinforcement learning tasks.
Reward function	In reinforcement learning, a function that provides feedback on the quality of the actions taken by a model, guiding the learning process by indicating which actions lead to higher rewards.