Term	Definition
Reward-weighted distribution	A probability distribution that has been adjusted based on the rewards obtained, used in reinforcement learning to guide the optimization of policies toward actions that yield higher rewards.
Rollout	The process by which a model generates different responses for a given query, used in reinforcement learning to evaluate the effectiveness of policies. In libraries like Hugging Face, the term can differ slightly but generally refers to the multiple possible outputs generated by a model.
Sampling	A technique used in language models where a model generates responses based on a probability distribution, selecting tokens randomly according to their probabilities.
Sentiment analysis pipeline	A sequence of processing steps in Hugging Face that evaluates the sentiment (positive or negative) of text, often used to score the quality of generated responses in models like chatbots.
Sentiment score	A score that reflects the sentiment (positive or negative) of a generated response, often used in training models like PPO to encourage the generation of responses with a desired sentiment.
Sigmoid function	A mathematical function that produces an S-shaped curve, commonly used in machine learning as an activation function or in logistic regression to map predictions to probabilities.
Softmax function	A function used in machine learning to convert the output of a model into a probability distribution, often applied in the final layer of neural networks for classification tasks.
Stats_all	A list or storage that holds the training statistics for each batch in a proximal policy optimization (PPO) training session, used to track performance metrics.
Stochastic gradient ascent (SGA)	An optimization algorithm that iteratively updates parameters to maximize a function, particularly useful in reinforcement learning where the objective is to maximize expected rewards.
Temperature (τ)	A hyperparameter in the softmax function that controls the randomness of predictions by scaling the logits before applying softmax. Lower temperatures make the distribution sharper, while higher temperatures make it more uniform.
Top-k sampling	A method in language models that restricts the selection of the next token to the top-k highest probability tokens, ensuring more focused and coherent outputs by filtering out less likely options.
Top-p sampling	A sampling technique where the model selects the next token from the smallest set of tokens whose cumulative probability exceeds a threshold p, allowing for more dynamic and context-dependent sampling compared to top-k sampling.
Trainer.train()	A method in Hugging Face Trainer class that initiates the training process of a model using the specified dataset and training arguments.
trainer.state	A property in Hugging Face's Trainer class that stores the state of the training process, including training logs, which can be used to monitor the progress and performance of the model during training.