Glossary: Generative AI Advance Fine-Tuning for LLMs

Welcome! This alphabetized glossary contains many terms used in this course. Understanding these terms is essential when working in the industry, participating in user groups, and participating in other certificate programs.

Term	**Definition**
Argmax/Argmin	The argument of the maximum/minimum of a function, i.e., the value of the input variable that yields the highest/lowest output value.
AutoModelForCausalLM	A class from the Hugging Face Transformers library used to load pre-trained causal language models like GPT-2 for text generation tasks.
AutoModelForCausalLMWithValueHead	An extension of the AutoModelForCausalLM class in Hugging Face for reinforcement learning, including a value head used for estimating the value function, crucial for models like PPO.
Beam search	A search algorithm that expands the most promising sequences of tokens at each step in sequence generation, used to improve the quality of outputs in language models by considering multiple possibilities simultaneously.
Beta (β) parameter	A hyperparameter in reinforcement learning that controls the balance between the current policy and the reference model, impacting the exploration and exploitation trade-off in policy optimization. In the context of DPO, it acts as the temperature parameter for the DPO loss.
Bradley-Terry model	A probabilistic model used for ranking and comparing different items or choices, often used to model pairwise preferences, where the probability of one item being preferred over another is based on their respective scores.
Closed-form solution	An explicit analytical expression for the solution of a problem that does not require iterative or numerical methods to solve.
Collator function	A function that organizes and batches input data into a format suitable for processing by machine learning models, especially in reinforcement learning scenarios.
Cost function	A function that represents the cost associated with a specific set of parameters in an optimization problem. It is used to guide models toward better performance by minimizing the cost during training.
Data collection	The process of gathering and preparing datasets, particularly preference datasets, for use in training models like those using direct preference optimization (DPO).
Dataset	A collection of data used for training, validating, and testing machine learning models. In this context, it refers specifically to the IMDB dataset used for sentiment analysis.
Dataset tokenization	The process of converting raw text data into token IDs that can be processed by machine learning models, particularly language models.
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.
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.
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.