

Machine Learning Operations (MLOps) with Vertex AI: Model Evaluation - Module 2

Find links to the resources and further reading mentioned in the training.

[Tackling Bias in Pre-trained Language Models: Current Trends and Under-represented Societies](#)

[Inadequacies of Large Language Model Benchmarks in the Era of Generative Artificial Intelligence](#)

[BLEURT: Learning Robust Metrics for Text Generation](#)

[Evaluation methods and metrics](#)

[Evaluation methods and metrics: Summarization](#)

[Evaluation methods and metrics: Question answering](#)

[Evaluation methods and metrics: Tool Use](#)

[Evaluation methods and metrics: General Text Generation](#)

[Self-Consistency Improves Chain of Thought Reasoning in Language Models](#)

[Supported tasks and criteria](#)

[Evaluating LLM Applications](#)