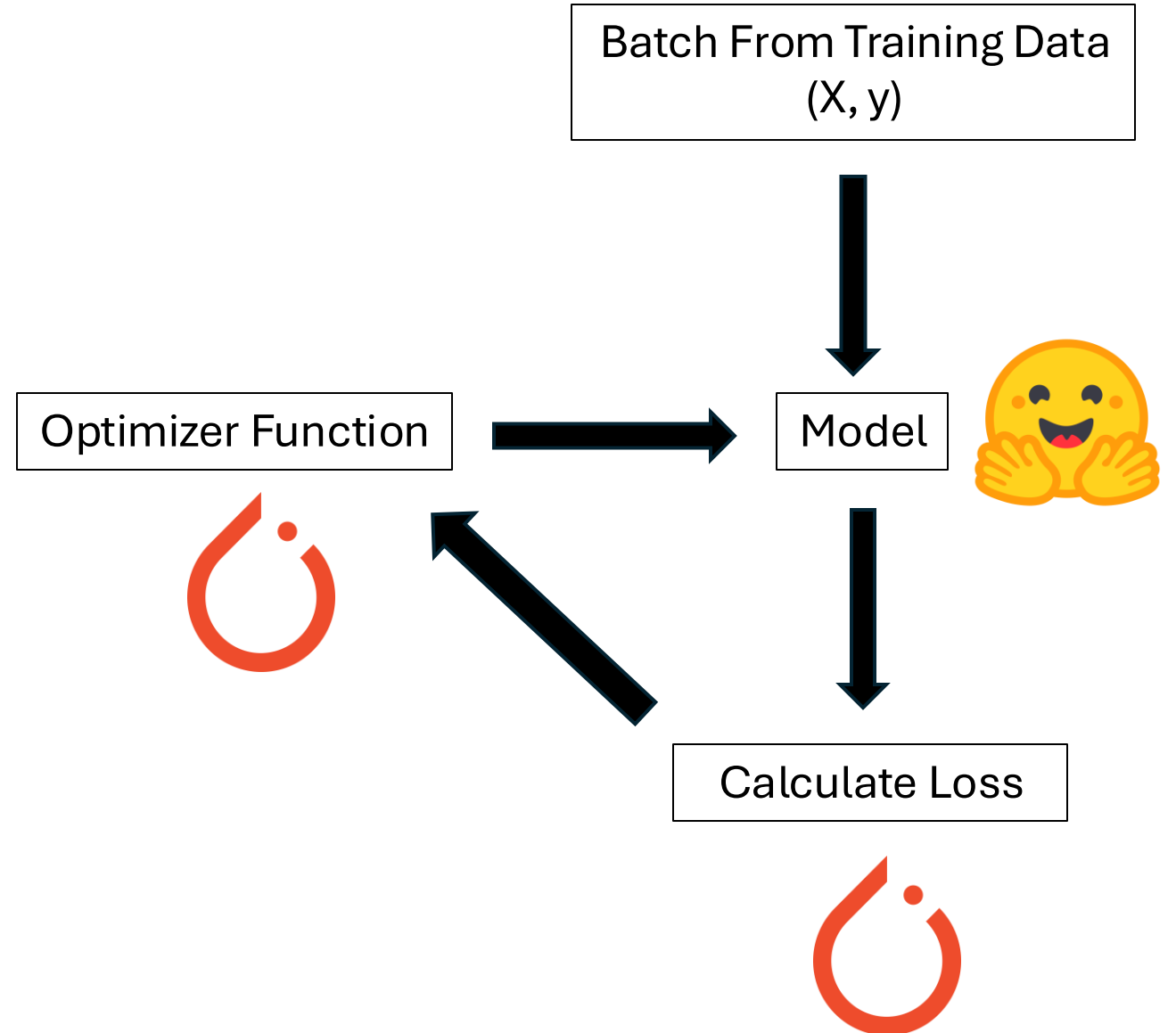


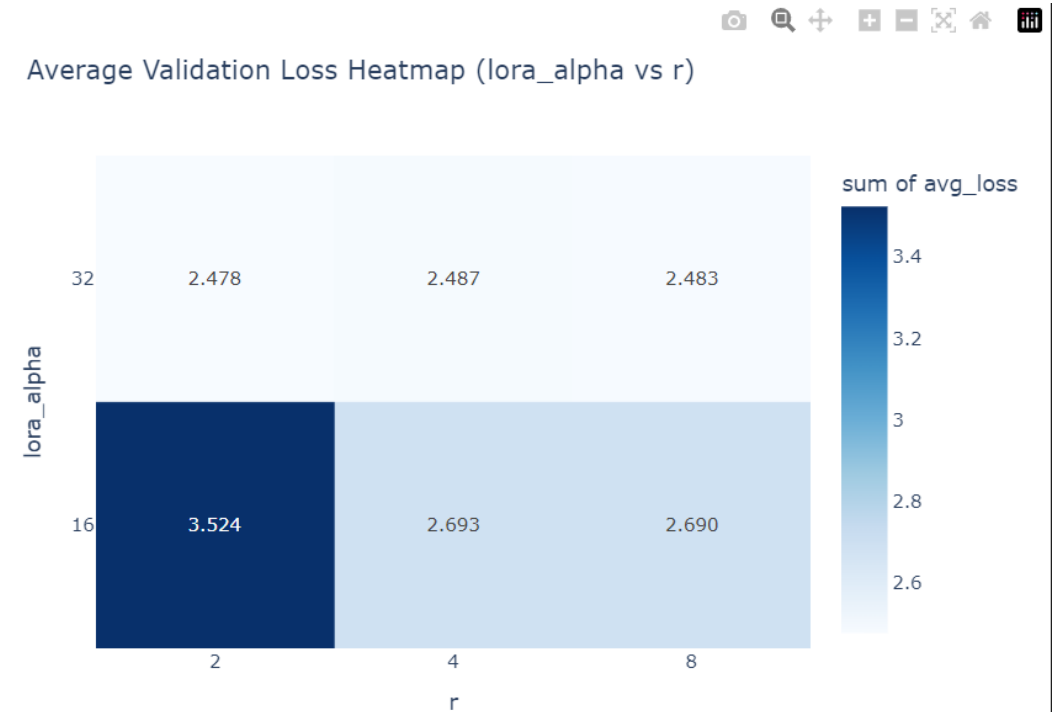
Fine-Tuning – Open Source Models



Grid Search Cross Validation

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
First Iteration	Test				
Second Iteration		Test			
Third Iteration			Test		
Fourth Iteration				Test	
Fifth Iteration					Test

Average Validation Loss Heatmap (lora_alpha vs r)

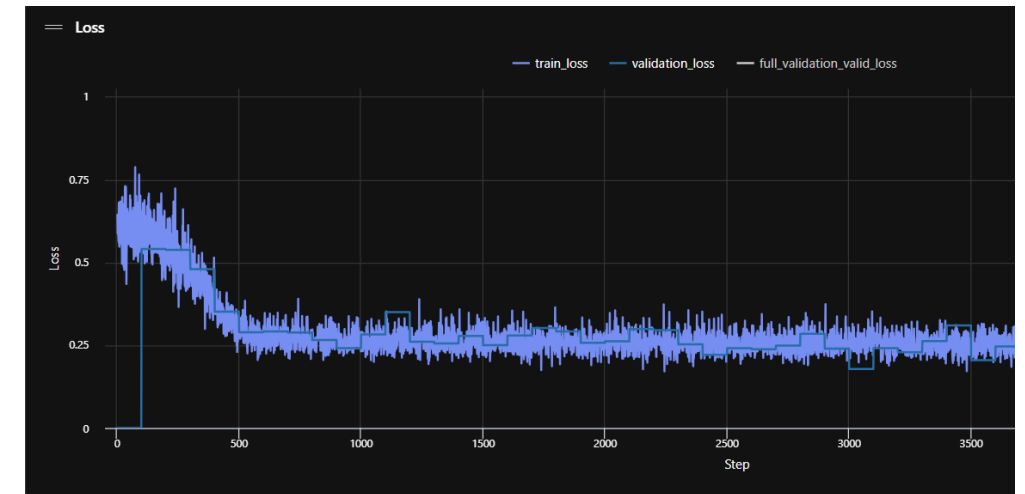


RAG Pipeline + Multi-shot Prompt Injection



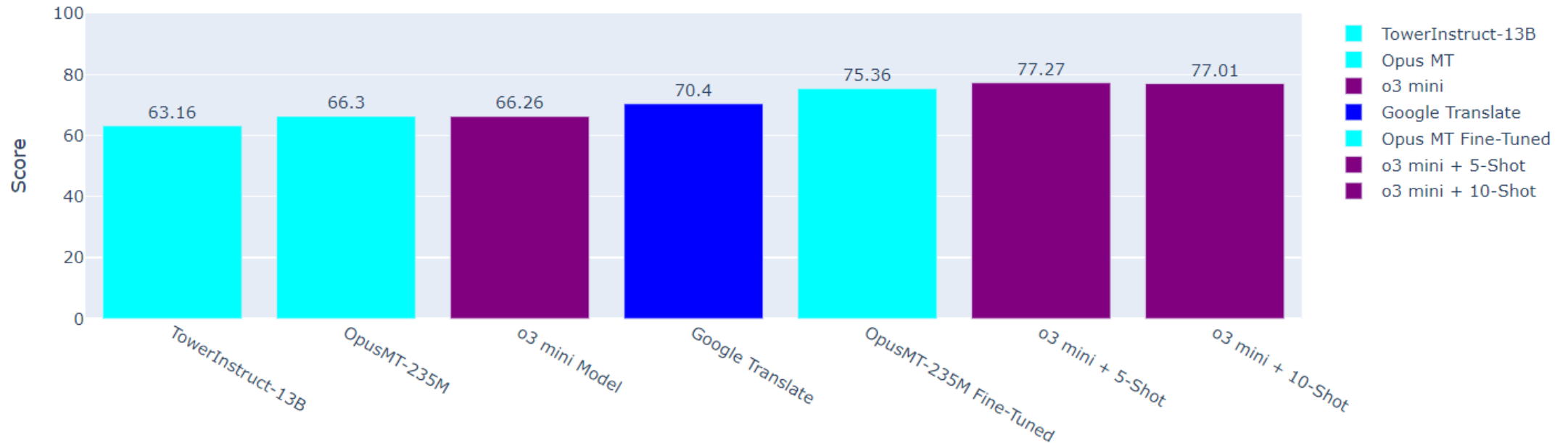
Fine-Tuning – Closed Source LLMs

Model attributes		Task parameters
ID	ftjob-dc179d0cb4b44ccbbc24b45ec9460ae6	Batch size
Status	▶ Running	32
Created on	Jun 11, 2025 11:18 AM	Learning rate multiplier
Validation file	val_abstracts.jsonl	0.01
Method of Customization	Supervised	Number of epochs
Auto-deployment enabled	true	3
		Seed
		42
Base model		
gpt-4.1-2025-04-14		
Training file		
train_abstracts.jsonl		
Azure OpenAI resource		
sahil-k-test-resource		
Training type		
Global (Preview)		
Deployment type		
Developer (Preview)		
Download results		Download training file
		Download validation file

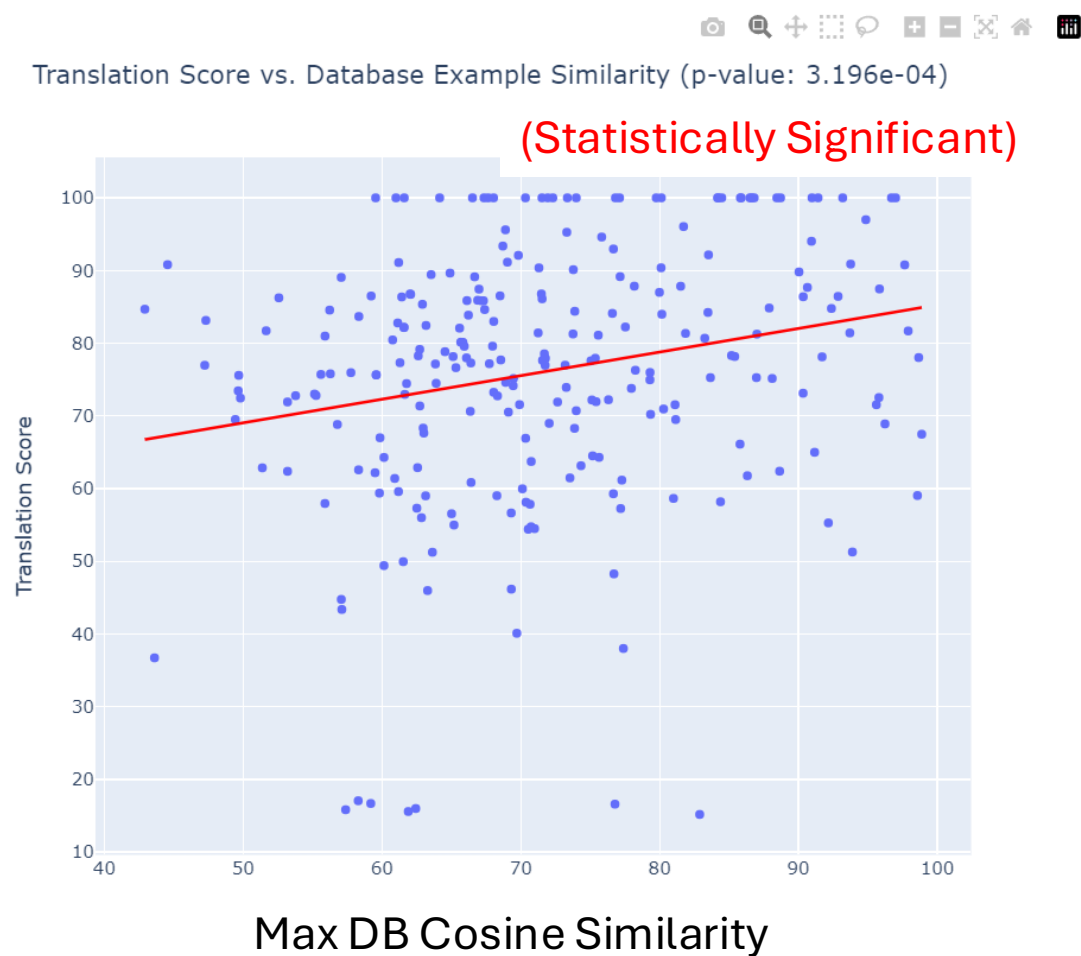
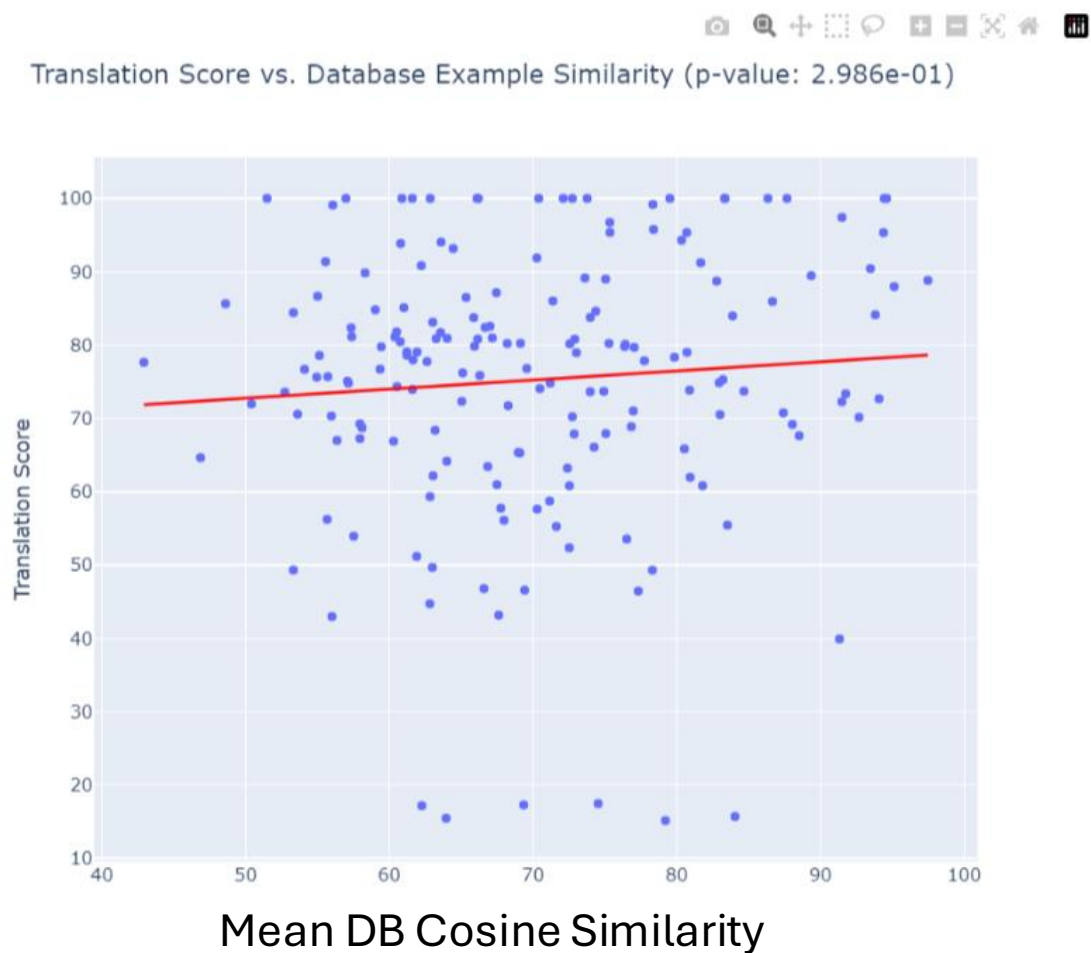


- **Metric: Weighted Combination of Rouge-L, Bert – Score, chr-F (Structure, Semantic, Character-level similarity)**

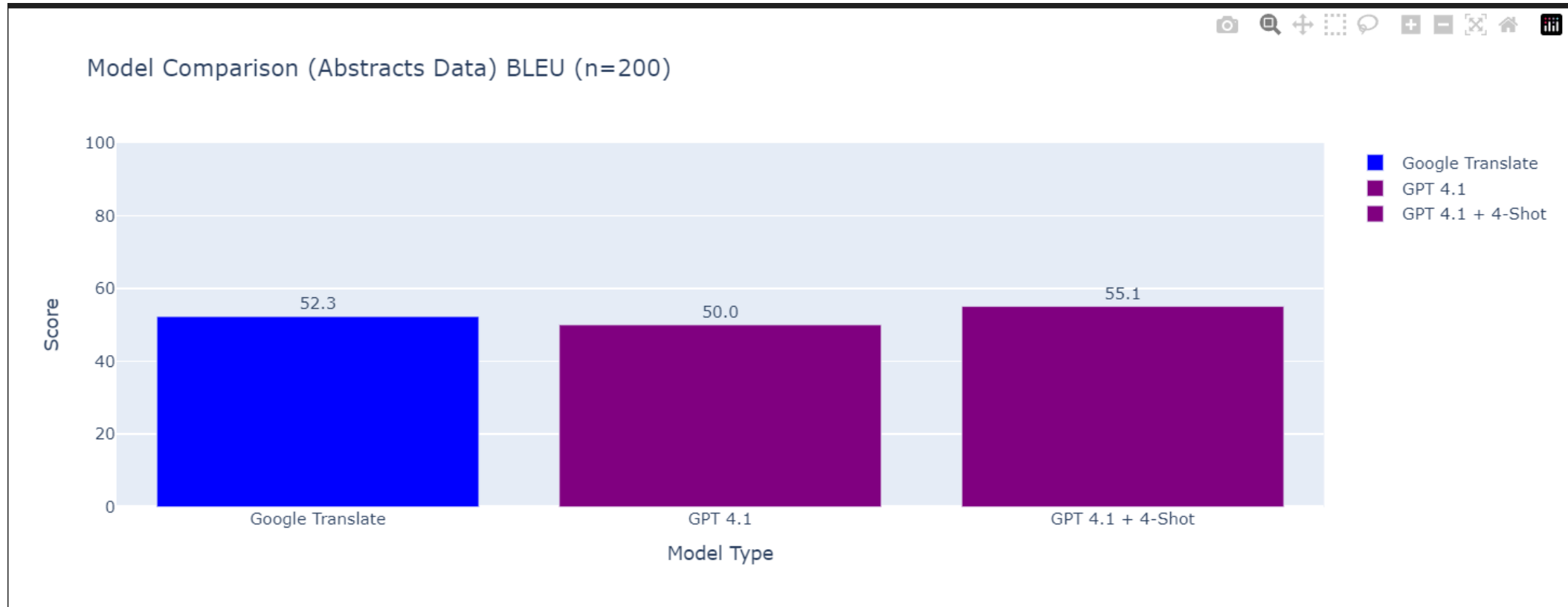
Model Comparison (Title Data) n=250



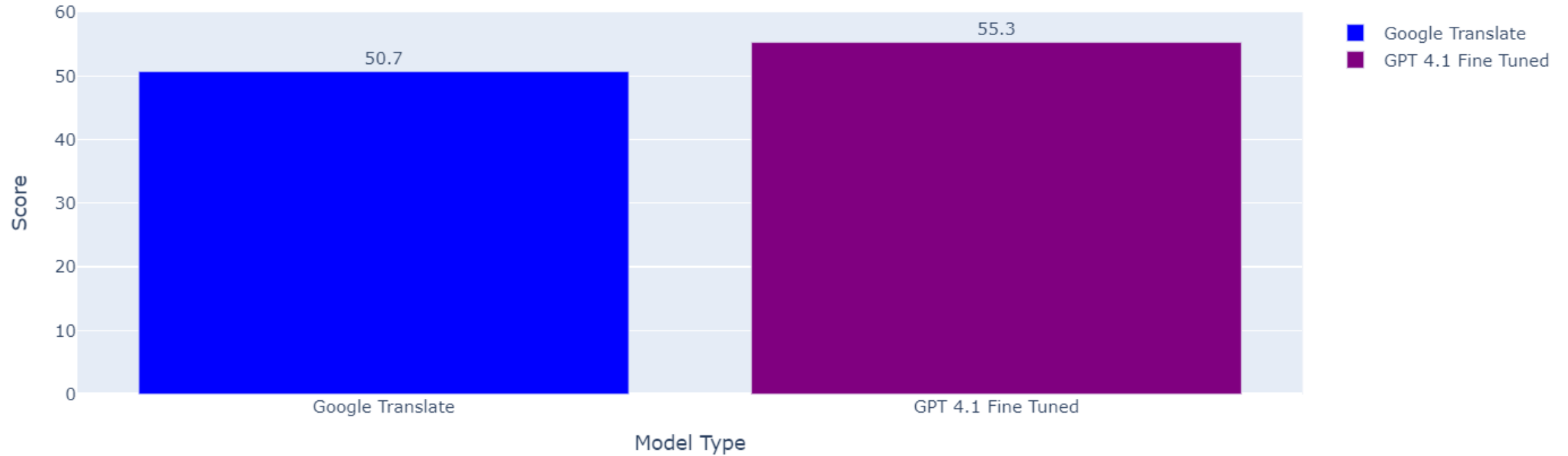
On the Vector DB



- **Metric: BLEU (Used by Statscan & NRC)**



Model Comparison (Abstracts Data) BLEU (n=2000)



- Metric: BLEU (Used by Statscan & NRC)