

W9

- Durchführung mehrerer Fine-Tuning-Runs mit neuen Adaptern auf RunPod.io.
- Verbesserung der Prompt-Templates und Optimierung der Inferenzparameter (Kontextlänge, Batchgröße).
- Erstes funktionierendes FastAPI-Backend mit dem Inferenzserver und stabiler Anbindung an das Plugin.
- Entwicklung einer initialen Version der Logik-Pipeline, um Modell-Outputs zu bewerten.
- Erweiterung und Bereinigung des Trainingsdatensatzes, um die Modellqualität weiter zu steigern.