

INSTRUCTION TUNING AI MODEL FOR CUSTOM PERFORMANCE



GUIDE TO INSTRUCTION TUNING PROCESS



CHALLENGE SOLVED : CUSTOMISED MODEL

- 1. OPEN SOURCE MODEL TRAINED ON GENERAL DATA WILL NOT PERFORM WELL IN SPECIFIC ENVIRONMENT/ TASKS**
- 2. MODELS CAN BE RE-TRAINED WITH NEW DATASETS WHICH IS THE FORM OF INSTRUCTION AND CORRESPONDING REPLY**
- 3. GENERATING SUCH DATASET CAN BE TIME CONSUMING, WHICH IS AUTOMATED USING TXTINSTRUCT**
- 4. ONCE GENERATED TRAINING THE MODEL WILL REQUIRE POWER, INFRA INVESTMENT, AND TIME TO TRAIN**

GENERAL STEPS TO INSTRUCTION TUNING

- 1.DECIDE ON THE BASE MODEL TO INSTRUCTION TUNE**
- 2.DECIDE ON THE TYPE OF INSTRUCTION THE MODEL NEEDS
TO BE TUNED**
- 3.DECIDE ON THE CONTEXT AROUND WHICH THE INSTRUCTION
AND CORRESPONDING REPLIES WILL BE GENERATED**
- 4.CURATE THE CONTEXT AND CREATE A BASE DATASET**
- 5.USE EXISTING TOOLS TO CREATE INSTRUCTION + REPLY
PAIR DATASET**
- 6.USE THAT DATASET TO TUNE THE MODEL**

TX TINSTRUCT DATASET CREATION

CONTAINS AN LLM
LIKE GPT / FLAN

**STATEMENT
GENERATOR**

GENERATE SOURCE
STATEMENT

THE STATEMENT THAT
FINAL MODEL WILL
NEED TO ANSWER

CONTAINS DATA /
EMBEDDING OF
CONTEXT

**KNOWLEDGE SOURCE
FOR CONTEXT**

SOURCE STATEMENT IS
GENERATED BY THE LLM
THAT READS THE
CONTEXT PROVIDED TO IT

CONTAINS AN LLM
LIKE GPT / FLAN

**LLM TRANSLATES
SOURCE TO TARGET
STATEMENT**

BASED ON THE TEMPLATES
/ PROMPTS THE TARGET
STATEMENT IS GENERATED

**STMT GEN +
KNOL SRC**

**DATASET
BUILDER**

INSTRUCTOR

EXTRACTOR

LETS PRACTICE

- THE NOTEBOOK IS UPDATED FOR EASIER PRACTICE
- USING OWN EMBEDDINGS IS MADE SIMPLE
- LEARNING THE CONCEPT OF INSTRUCTING THE MODELS REQUIRES VERY LESS TIME
- DATASETS AVAILABLE ELSEWHERE CAN BE EASILY PORTED TO THIS NOTEBOOK
- WORKING WITH DIFFERENT BASE MODELS CAN BE A GOOD EXERCISE, IF YOU HAVE SUFFICIENTLY LARGE WORKSTATION
- MAKING DATASETS IS THE KEY ASPECT OF THE NOTEBOOK.

insightbuilder/ python_de_learners_data



Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

2
Contributors

0
Issues

40
Stars

19
Forks

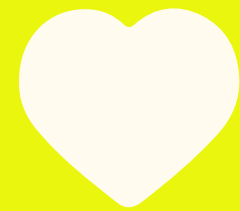


python_de_learners_data/txtInstruct_trial_ver03.ipynb at main · insightbuilder/python_de_learners_data

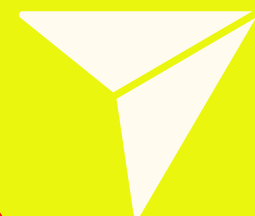
Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python_de_learners_data/txtInstruct_trial_ver03.ipynb at main...

GitHub

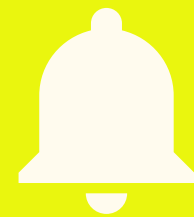
THANKS FOR WATCHING



LIKE



SHARE



SUBSCRIBE