Llm coding   
**Option 1: Manually Label a Subset of Your Dataset**

This is ideal for your context-aware sentiment analysis project.

**Steps:**

1. **Extract a subset** of samples from frankdarkluo/dailydialog:

python

CopyEdit

from datasets import load\_dataset

dataset = load\_dataset("frankdarkluo/dailydialog", split="train")

samples = dataset.select(range(100)) # Choose first 100 for example

1. **Export to CSV for manual annotation**:

python

CopyEdit

import pandas as pd

df = pd.DataFrame(samples)

df["manual\_label"] = "" # Leave empty for now

df.to\_csv("samples\_to\_label.csv", index=False)

1. **Open the CSV in Excel or Google Sheets**, read the context + response, and manually assign labels:
   * Positive, Negative, Neutral
2. **Load labeled data back for use**:

python

CopyEdit

df = pd.read\_csv("manually\_labeled.csv")

from datasets import Dataset

labeled\_dataset = Dataset.from\_pandas(df)

Now you can use labeled\_dataset to:

* Fine-tune a model (like RoBERTa)
* Evaluate other models against true labels

| **Aspect** | **daily\_dialog (original)** | **frankdarkluo/dailydialog** |
| --- | --- | --- |
| Format | Full dialogue as list of utterances | Flattened into context-response pairs |
| Labels | Emotion + Dialogue Act labels | ❌ No emotion or act labels |
| Preprocessing | Structured, original format | Preprocessed (simplified, cleaner) |
| Availability | Gated initially | Publicly accessible on Hugging Face |
| Good for | Fine-tuning + supervised tasks | Zero-shot + inference + weak supervision |