1> Dataset for latex code from natural language text
https://github.com/knowledge-verse-ai/TeXpert

2> LLM for converting plain text, word or pdf to Latex code (project not yet completed)
https://github.com/elotech47/doc2latex-llm

3> detects labels and symbols from P&ID diagrams
https://github.com/shailavij/P_ID-Symbol-Objectdetection

4> symbol detection and labelling
https://github.com/aneeshbhattacharya/Automated-PnID-Symbol-Detection-and-Labelling

5> AutomaTikZ
takes **natural-language descriptions** and generates TikZ code
It is basically finetuned llama
https://github.com/potamides/AutomaTikZ

6> DaTikzV2
https://huggingface.co/datasets/nllg/datikz-v2
DaTikZv2 is a dataset of TikZ drawings aligned with captions

7> DeTikzify
multimodal language model that automatically synthesizes scientific figures as semantics-preserving TikZ graphics programs based on sketches and existing figures.
https://github.com/potamides/DeTikZify

8> vTikZ benchmark

So we don't have any model for converting tikz code to natural language instructions
Task: `TikZ code → step-by-step natural language instructions`.

DaTikZ dataset ma we need to convert the captions into human readable instructions using some strong general purpose LLM GPT-4o or claude.

Then just finetune LLama on it and its done