This is the file to explain the code and dataset published on the github. Note that as for dataset annotation, all the instructions and possible utility have been given to the volunteers.

As for the *auto_detect_span.py*, this is used to preprocess the dataset with full NL-TL pairs and to process them into lifted NL-STL. The raw data are given in the data dir.

As for the *T5_tuning_liftTL.py*, this is to finetune the T5 model with lifted NL-TL dataset. The specific code is to directly run it: *python T5_tuning_liftTL.py*

As for *transfer_learning.py*, this is to do the further finetuning on the specific domain for full NL-TL transformation. There are some human-setting parameters to determine the task, like the dataset used, the initial weights (pre-trained or not pre-trained), and the model type (T5-large or T5-base). The example command is:

python self_tuning_model_different_training_size.py \
-seed 309 \
-name circuit \
-init_weight with_pre-train \
-data_size 0.01-0.09 \
-model_checkpoint t5-large

As for *transfer_learning_CW.py*, it is similar to *transfer_learning.py*, but focused on CW dataset.

As for *Seq2seq_lifted_all.py*, it is to train the Seq2Seq baseline, the dataset can be all the lifted NL-TL pairs or the full NL-TL pairs of each specific domain. The example command is:

python Seq2seq_lifted_all.py \
-seed 316 \
-data_size 0.01-0.09

For the dataset creation with GPT-3, there are two frameworks. However, you need to put your Openai Key into it. Before run the code, remember install Openai, *pip install openai*