What is SOTA in NLP?

This aims to track the progress in Natural Language Processing (NLP) and give an overview of the state-of-the-art (SOTA) across the most common NLP tasks and their corresponding datasets

**Dual Intent and Entity Transformer(DIET)** -a transformer architecture help intent classification and entity recognition together.

The best thing about DIET is its flexibility. It provides the ability to plug and play various pre-trained embeddings like BERT, GloVe, ConveRT, and so on.

So, based on your data and number of training examples, you can experiment with various SOTA NLU pipelines without even writing a single line of code.

In Diet :- **total loss** = **entity loss + mask loss +** **intent loss**

The only files that we care about for this tutorial are:**/data/nlu.md-**This markdown file will hold our training data.

* **config.yml-**This file will be used to define out NLU pipeline.

The steps are as follows:

• The message is received and passed to an Interpreter, which converts it into a dictionary including the original text, the intent, and any entities that were found. This part is handled by NLU.

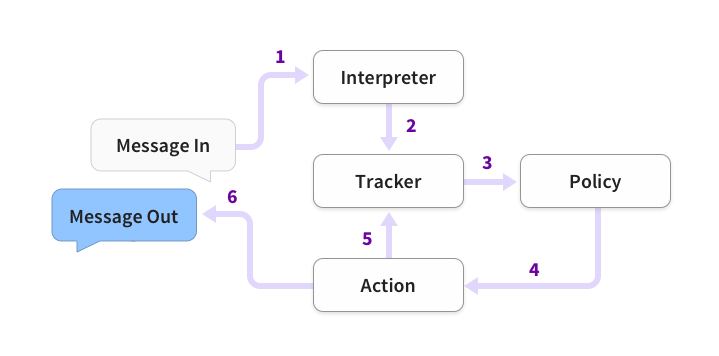
• The Tracker is the object which keeps track of conversation state.

• The policy receives the current state of the tracker.

• The policy chooses which action to take next.

• The chosen action is logged by the tracker.

• A response is sent to the user.



**Training Data : nlu.md**

To mark a word or group of words as an entity, enclose the words in [] followed by the name of the entity enclosed in ()

## intent:inform\_quota\_code  
- Quota is [General](quota\_code)

If any of the entities have a finite set of values then you can also add a lookup table for that entity. For example, the entity *country name*has a finite set of values.

## lookup:countries <!-- lookup table list -->  
- India  
- Nepal  
- China

## **NLU Pipeline :** config.yml

Two parts :-

1. The first one is for the **NLU pipeline**

2 . policies for Rasa Core.

Pipeline:

- name: WhitespaceTokenizer

- name: RegexFeaturizer

- name: LexicalSyntacticFeaturizer

- name: CountVectorsFeaturizer

analyzer: "char\_wb"

min\_ngram: 1

max\_ngram: 4

- name: DIETClassifier

epochs: 100

- name: EntitySynonymMapper

- name: ResponseSelector

epochs: 100

Ordering of the components of the NLU pipeline is also very important. For example, you can’t define featurizers before the tokenizers because the output of the preceding component acts as the input for the following component.