**Download anaconda:**

* <https://www.anaconda.com/products/distribution>
* create new environment : anaconda à environment à create à enter name and choose the python version you want

**Download spacy :**

* by running the commands given in the link <https://spacy.io/usage/>

in powershell prompt (anaconda à home à install powershell prompt and launch) after choosing the suitable options

**\*make sure to choose your environment in (home)**

**\*if you are using a device with graphic card make sure to choose hardware à GPU (CUDA version same as the device supported CUDA version)**

**\*to know the CUDA version supported run the following command in command prompt :**

**nvcc –version**

**Download prodigy :**

* **To download prodigy:** [Installation & Setup · Prodigy · An annotation tool for AI, Machine Learning & NLP](https://prodi.gy/docs/install) **\* if you faced (filename.whl is not a supported wheel on this platform) error à Make sure that the supported python version of .whl file is suitable for the python version installed (e.g. cp38 à python 3.8)**
* **To use prodigy**: [GitHub - explosion/jupyterlab-prodigy: 🧬 A JupyterLab extension for annotating data with Prodigy](https://github.com/explosion/jupyterlab-prodigy)

**For each jsonl file:**

1. !python -m prodigy db-in (name of db) (jsonl file name/path) **----to create dataset for jsonl file**

2. !python -m prodigy data-to-spacy output\_dir --ner --textcat --textcat-multilabel --tagger --parser --senter --spancat --eval-split --config --base-model --lang --verbose -F **---to convert from jsonl to spacy format**

3. add paths for train and dev in cfg file

4. !python -m spacy train corpus\config.cfg --paths.train corpus\train.spacy --paths.dev corpus\dev.spacy **----train data on config**

**For BERT Models Train:**

1. Download spacy transformers : pip install spacy-transformers
2. Create config file with transformers : [Training Pipelines & Models · spaCy Usage Documentation](https://spacy.io/usage/training#config)
3. Change the model name in the config file + add paths for train and dev (after fill-config)
4. debug config file to make sure nothing is missed using this command :

!python -m spacy debug data NER\config\_ner.cfg

5. Train : !python -m spacy train corpus\config.cfg --paths.train corpus\train.spacy --paths.dev corpus\dev.spacy

**For FastText Model Train:**

1. Download language vec file from : <https://fasttext.cc/docs/en/crawl-vectors.html>
2. Run the following command: !python -m spacy init vectors ar POS/cc.ar.300.vec /ft\_vectors
3. use the path of the vectrors in your config as the 'vectors' setting in [initialize].
4. train on config file

**Commands:**

* **To debug config file :** !python -m spacy debug data NER\config\_ner.cfg
* **To fill config file :** !python -m spacy init fill-config POS\POSconfig.cfg POS\POS\_config.cfg
* **To use prodigy for annotation :** !python -m prodigy pos.correct Batch-1+2+3+4.pos blank:ar ./Batch-1+2+3+4.pos.jsonl --label ADJ,PUNCT,NOUN