Class Finetune\_Dataset(Dataset)

Třída bude sloužit ke zpracování dat pro Finetuning model

Metody:

1. Def \_\_init\_\_

* Dědí init z Dataset torch.utils.data
* Zpřistupňuje proměnné dataset , instanci tokenizeru a max seq lengt povolenou modelem aby se s nimi pracovali uvnitř třídy

1. \_\_len\_\_

* Aby šlo na objekt zavolat len pro zjištění délky datasetu (standardní věci)

1. \_\_getitem\_\_

* Také standardní věc při zpracování datasetu, díky které po vytvoření instance třídy můžeme pracovat s objektem prakticky jako s array , takže můžeme přistupovat k jednotlivým záznamum k trenování (jednotlivých batchi)
* Zde dochází následně k předzpracování dat tak, že celý zánam převedeme na jednu dlouho sequency tokenu
* Zaroveň vytvoříme zde masku , jelikož tokeny pro assistanta chceme umět maskovat, jelikož to bude to co model bude předpovídat