Benchmark etnograficzny – kod Natalii  
  
1. Parser i podawanie parametrów do skryptu – u Natalii w osobnym JSON, który trzeba zmienić za każdym razem, ale main uruchamia się bez dodawania ich od myślników  
2. Natalia ma klasy dla każdego modelu (plik models.py)

3. Folder z wynikami, gdzie byłby plik dla każdego modelu – lista pytań, jak odpowiedział model i czy to było poprawne

4. Osobny plik JSON z podsumowaniem – liczba pytań, liczba poprawnych, liczba niepoprawnych (ale żeby pytania nie były widoczne), chyba dla wszystkich modeli jeden.

5. Lista odpowiedzi to ma być lista JSONÓW (jak w pdf) a podsumowanie osobny plik? Dopytać.

Aktualizacja 30.07.2025:

Sprawdzić czy z modelami z Huggingface można się łączyć przez API zamiast pobierać lokalnie.   
Przetestować jakiś model żeby sprawdzić jak wyglądają odpowiedzi.  
Przy API sprawdzić OpenAI bo może być jakiś problem…  
  
Na 13. Sierpnia się umawiamy.  
TODO: dodać obsługę API dla pierwszej części skryptu – gogle, openai i może vllm(?)  
Zrobić plik testowy z kilkoma pytaniami – żeby były różne dziedziny.  
Przetestować całość na modelu lokalnym i przez API.