

# Projekt

Deklaracja tematu projektu i planowanego do wykorzystania zbioru danych – **do 10 kwietnia** – w Zadaniu (Assignment) w MS Teams.

**Realizacja projektu: do 30 maja**

**Implementacja/kod (8 punktów) + sprawozdanie (6 punktów) – łącznie 14 punktów**

**Proponowane zadanie** – ale możliwe jest zaproponowanie własnego tematu projektu; bardziej złożone projekty mogą być realizowane w parach (po uzgodnieniu)

1. **Klasyfikacji tekstu na wybranym zbiorze danych** (np. detekcja spamu, detekcja fałszywych wiadomości, klasyfikacja tematu/typu dokumentu)
  - a. Wybierz zbiór danych
    - i. Wiele zbiorów danych dostępnych jest w serwisie Hugging Face, np. zbiory danych dotyczące detekcji spamu czy fałszywych wiadomości (fake news).
  - b. Zaimplementowanie i eksperymentalne porównanie skuteczności **dwóch podejść (modelu generatywnego i wybranego innego)**:
    - i. Ekstrakcja cech z tekstu metodami klasycznymi, np. TF-IDF lub BM25. Metoda uczenia maszynowego oparta o tak wyznaczone reprezentacje uczona na zbiorze treningowym.
    - ii. Klasyfikator probabilistyczny oparty o wybrany model klasy tylko-koder dostrojony na zbiorze treningowym.
    - iii. **Dostosowanie modelu generatywnego** (klasy tylko-dekoder) do generowania odpowiedzi w oczekiwanym formacie. Np. w klasyfikatorze spamu generowania „tak” lub „nie”. Sprawdzenie metody in-context learning i wybranej metody (lub metod) efektywnego dostrajania modelu (metody rodziny PeFT), np. soft-prompt tuning czy LoRA.
  - c. Ewaluacja każdego podejścia
    - i. Wyznaczenie metryk takich, jak dokładność, precyzja i czułość dla każdej klasy, macierz pomyłek na zbiorze testowym.
    - ii. Wizualizacja wyznaczonych reprezentacji (dla części zbioru testowego) metodą tSNE lub UMAP.

**Dodatkowe informacje:**

- W przypadku niewystarczających zasobów obliczeniowych można ograniczyć licznosc zbioru danych, albo skrócić przetwarzane dokumenty.
- Przykładowe zbiory danych z Hugging Face do trenowania modelu klasyfikacji fałszywych wiadomości:
  - ErfanMoosaviMonazzah/fake-news-detection-dataset-English
  - GonzaloA/fake\_news
- Lekkie modele klasy tylko-koder
  - DistilBERT <https://huggingface.co/distilbert>
- Wykorzystać lekkie modele generatywne z biblioteki Hugging Face
  - Modele z rodziny SmoLM2: <https://huggingface.co/HuggingFaceTB>
    - Modele z rodziny Phi, np Phi 1.5 – 1.3B parametrów ([https://huggingface.co/microsoft/phi-1\\_5](https://huggingface.co/microsoft/phi-1_5)) czy Phi 4-mini – 4B parametrów
  - Polskie modele językowe z rodziny Qra (Qra-1b) <https://huggingface.co/OPI-PG/Qra-1b>