Language Models
Lab 1
Sessions 1 and 2
Note: You can transfer 2 tasks to Lab 2 without losing points.

On the SKOS, you will find a link to the materials needed to solve the tasks from this list. In each task, you should use a language model for the Polish language from Hugging Face (English replacements are also allowed: instead of papuga use GPT-2 (version with approx. 100M parameters), instead of polka use a variant of LLama 1b.

Zadanie 1. (**4p**) Using a selected language model, create a chatbot (you have full freedom regarding what the chatbot should talk about; it doesn't need to perform any useful function). Your chatbot should:

a) Generate reasonably short, somewhat meaningful responses.

b) Manage the dialogue history in some way, partially including it in the prompt.

c) Generate more than one response and choose the óptimalżesponse based on some criterion you select.

We'll likely revisit this task, so don't worry if the result is not fully satisfactory for now (it can be challenging to prompt some models trained on texts to engage in dialogue).

Zadanie 2. Consider the following three sentences:

> John likes spinach very much.
> Squirrels live in the park.
> Last night, I met a wonderful woman who passionately talked about language models.

Your task is to write a program that, given a (multi)set of words, outputs several example arrangements of those words, ordered from the most natural. Additionally, we assume that the period is always at the end, and the only word capitalized is the first word of the sentence. The **only** source of knowledge about the language should be a language model that can run on your available computer.

The task has two separately graded variants:

a) The number of words is small enough to go through all permutations. (**3p**)

b) The number of words is small enough to examine all pairs, find words that clearly go well together, form probable word chunks, and permute only those. (**4p**)

Test your program(s) on the provided example sentences and a few of your own.

Zadanie 3. (**4p**) Using a language model and functions for assessing text probability, write a program that detects the sentiment of a review (whether it is positive or negative). Example positive reviews:

> Monitored parking included in the price.
> Clean hotel, rooms were thoroughly cleaned.
> Overall, I can recommend him, he guided me through the necessary tests, analyzed the results, and patiently answered questions.
> Great post-industrial vibe in the tenement houses.
> Playroom for kids, playground outside, fireplace, table tennis.

Example negative reviews:

Air conditioning not working in many rooms.
Despite the so-called European days, the food was monotonous.
More expensive than competitors with similar standards.
Maybe I gave up too soon, but I didn't want to spend money on meetings that weren't yielding results.
Avoid this place!

In this task, **do not** use few-shot learning techniques, only operate on the probabilities of certain phrases. Try more than one method of constructing these phrases. Provide the accuracy of your program (how often you find the correct answers) for the data from the lecture website (a representative random sample is sufficient).

Zadanie 4. (**6p**) Write a program that answers simple factual questions. The questions are varied, but you'll notice certain groups of them (sorting helps). They are taken from the competition https://2021.poleval.pl/, but you shouldn't look there (yet). The basic technique will be zero-shot/one-shot/few-shot learning and the polka model, but additionally, you should:

1. Handle at least one group of questions with a heuristic.
2. In at least one group of questions, use a language model that operates in probability estimation mode.

You are not allowed to use any form of information retrieval; the only source of knowledge should be the language model and the questions in the learning portion of the dataset. Provide the (estimated) percentage of correct answers. In the aforementioned competition, the baseline solution provided by the organizers had a few percent accuracy – it should not be hard to beat that.