Modele językowe
Ćwiczenia 1
Zajęcia 3

Zadanie 1. Come up with a type of task that is relatively easy for humans (not relying on factual knowledge), with a clearly defined answer,[1] that ChatGPT struggles with (alternatively: show three non-trivial tasks that Chat can do, which somewhat surprised you).

Zadanie 2. Choose 10 rather challenging questions from Task 4 in Workshop 1, and then check how many of them are correctly solved by ChatGPT. You can also decide to use other trivia questions.

Zadanie 3. In the case of language models, a watermark is a property assigned to generated texts that is invisible without specialized watermark detection methods. A certain language model provider proposed the following solution: add a condition (without notifying anyone) to prefer words starting with the letters V, S, and K during generation. Expand on this idea by considering how (and if) such a watermark would work for long texts. And for short texts?

Zadanie 4. At the Artificial Intelligence Olympiad, there was a task called Riddles[2]. In summary, it involved matching words to texts describing those words, with answers limited to a known set (containing about 10,000 words).

An example text was:

> a woman traveling by a mode of transportation, e.g., plane, train, ship

where the answer is *passenger*.

Participants in the Olympiad had access to word definitions from Wiktionary, but we will assume that we only have access to a language model. Answer the following questions:

a) Can any of our models be forced to solve this task in generation mode? What strategy do you think is promising?

b) Can the sentence probability scoring function be successfully used in this task? (what is the main problem with using it?)

Zadanie 5. Describe a procedure that, using a basic language model text generation interface, solves the problem of generating **exactly** one word (for a given prefix consisting of whole words) as efficiently as possible.

Zadanie 6. Read and discuss the biases of the papuGaPT model. How were these biases studied? What conclusions can be drawn from these studies? (section Bias Analysis on the page `https://huggingface.co/flax-community/papuGaPT2`).

Zadanie 7. It turns out that the Polka[3] model can translate simple English sentences into Polish. Show how a specific prompting technique can yield this capability. Provide two example sentences: one correctly translated by Polka and one with errors.

Assuming we have access to an English-Polish dictionary, propose a way to use this access to improve translation quality.

Zadanie 8. Assume we have two language models and want to generate text using knowledge from both. Propose 3 different scenarios, with at least one that does **not** assume the same tokenization in both models.

Zadanie 9. Propose a solution, different from the one suggested in the workshop list, to the word permutation selection task that partially mitigates issues related to greedy word pairing. The solution should use language models (for texts of varying lengths) and should not examine all permutations.

[1] Complex philosophical questions are excluded; the answer must be simple and indisputable

[2] `https://github.com/OlimpiadaAI/I-OlimpiadaAI/blob/main/first_stage/riddles/zagadki.ipynb`

[3] you can use also Llama-1B or compatible