**Home Assignment 2: Deep Learning and Embedding (5 points total)**

Aim of this assignment is to give you a chance to practice some basics of dimensionality reduction techniques from a perspective of deep learning. Your assignment consists of two tasks: essay and practical part.

**Essay (3 points)**
Your task is to read one paper from the list below and write an essay on it. In the essay you will have to explain in your own words and your own examples provide a brief explanation of the key aspects of the paper (minimum 2 pages).

The list of papers:

- Neural Networks Fail to Learn Periodic Functions and How to Fix It [link],
- How to avoid machine learning pitfalls: a guide for academic researchers [link],
- Billion-scale similarity search with GPUs [link],
- Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks [link],
- All Bark and No Bite: Rogue Dimensions in Transformer Language Models Obscure Representational Quality [link];

**Practice (2 points)**
As the practical part of the assignment you will have to practice similarity search and dimensionality reduction tasks. What you will have to do:

Part 1: Create a dataset from the paper you choose for the essay. Each object of data has to be a complete sentence.

Part 2: You should choose an open-source model, which you will use to generate embeddings of the text [https://huggingface.co/models].

Part 3: Generate embeddings of your dataset.

Part 4: Perform and evaluate any dimensionality reduction technique on your embeddings. Analyze components of your embeddings and use them to reduce an output of LLM.

Part 5: Reduce an output of LLM using information you retrieved from embeddings dimensionality reduction. (for instance, output size of embedding is reduced from vector of length 784 to vector of length 128).

Part 6: Create a search query. Conduct similarity search for created query. Analyze 10-closest results for both models, find intersection between 10-closest objects. Use string similarity methods (such as Levistein distance to evaluate results, as well).

Be aware:

- Do not forget to create a pivot table with results evaluation and visualize intersections between data objects;

**Submission:**

- Deadline: June 30, 23:59,
- Submit single .ipynb file;