PMLDL Progress Submission

Anton Kudryavtsev
Innopolis University
a.kudryavtsev@innopolis.university

Sofia Tkachenko
Innopolis University
s.tkachenko@innopolis.university

Anatoly Soldatov
Innopolis University
a.soldatov@innopolis.university

I. Entity Extraction

A. Project Topic

Named Entity Recognition (NER) is one of the most crucial subtasks in information retrieval systems and systems with large language models (LLM). NER algorithms are used to identify and classify key elements within text, such as names of people, locations, organizations, dates and other specific term and match them with corresponding pronouns (he, she, it, they). NER approaches have improved significantly in recent years due to the development of deep learning (DL) and natural language processing (NLP) techniques. This has been further enhanced by the creation of transformer models and large language models (LLMs). While generative LLMs, such as the generative pre-trained transformer (GPT), are gaining popularity in both the research and business worlds, NER is becoming increasingly important. This is because it can improve the accuracy of models, reduce bias, and protect personal data within a company's systems. Therefore, robust and scalable NER solutions would be beneficial for systems that include (NLP) or generative artificial intelligence (AI) approaches and work with personalized data or term saturated data.

B. Link:

https://github.com/PodYapolskiy/entity-extraction

C. Intermediate results

So far we have researched already existing solutions, as well as available python libraries. We have analyzed different complexities of implementation, and decided that this will be the most challenging and useful project for our team!

Also, we have preliminary defined the roles for our team:

- Anton - Backend and Infrastructure
- Sofia - Data preparation and Model development
- Anatoly - Research and Model development