Contribution ID: **1** Type: **Training**

Wykorzystanie dużych modeli językowych dla danych prywatnych - LLM for Private Data

Szkolenie dotyczy możliwości wykorzystania wielkich modeli językowych (ang. Large Language Models - LLM) do interakcji z prywatnymi danymi. Wprowadzone są w nim narzędzia, które pozwalają wykorzystać moc generatywnej sztucznej inteligencji w scenariuszach, w których dane nie mogą w żadnym momencie opuścić prywatnego środowiska.
Przedstawiona jest architektura i wymagania dotyczące danych w celu utworzenia prywatnego ChatGPT, przy wykorzystaniu zrozumienia semantyki zdefiniowanego problemu i jednoczesnego zachowania kontroli nad prywatnymi danymi.

Szkolenie obejmuje następujące tematy:
- Przegląd wielkich modeli językowych LLM
- Wprowadzenie do modeli językowych LLM i ich zdolności do generacji tekstu oraz przetwarzania języka naturalnego.
- Transformery
- Omówienie architektury transformerów, które stanowią podstawę działania modeli LLM.
- Przedstawienie mechanizmu uwagi (ang. attention) jako kluczowego elementu w modelach LLM i omówienie, jak mechanizm uwagi pozwala modelom LLM na selektywne koncentrowanie się na określonych informacjach.
- Przykłady zastosowania modelu uwagi w modelach LLM.
- Różne metody dostosowywanie modeli LLM
- Przedstawienie różnych technik używanych do dostosowywania LLM do konkretnych zadań, dziedzin i danych
- Dopasowywanie (ang. fine-tunning) modeli LLM
- Proces dalszego trenowania pretrenowanych modeli LLM na specyficznych danych.
- Wykorzystanie fine-tuningu do dostosowania modelu do konkretnych zastosowań (ćwiczenia praktyczne)
- Retrieval Augmented Generation (RAG)
- Omówienie podejścia RAG, które łączy generację tekstu z wyszukiwaniem informacji z wewnętrznego zbioru dokumentów
- Przykłady zastosowania RAG (ćwiczenia praktyczne)
- Przykłady zastosowania zaawansowanych technik poprawiających działanie RAG (ćwiczenia praktyczne)

Szkolenie zostanie przeprowadzone w języku polskim lub angielskim, w zależności od ewentualnego udziału uczestników niepolskojęzycznych.
Czas trwania: 6h (4h cz. 1 + 2h cz. 2)
Pierwsza część: środa, 2 kwietnia po południu (15:00-19:00)
Druga część: czwartek, 3 kwietnia (11:20-13:20)
Prowadzący:
- Rafał Możdzonek, z wykształcenia fizyk jądrowy, od wielu lat zajmuje się analizą danych, uczeniem maszynowym i pisaniem aplikacji do zastosowań naukowych wykorzystywanych w Narodowym Centrum Badań Jądrowych.
- dr Krzysztof Nawrocki, fizyk cząstek wysokiej energii w Narodowym Centrum Badań Jądrowych z dużym doświadczeniem w projektach z zakresu big data, grid computing, astronomii i uczenia maszynowego. Specjalizuje się w analizie danych tekstowych, szczególnie w dziedzinie medycyny.

Liczba uczestników jest ograniczona. Decyduje pierwszeństwo zapisu.

LLM for Private Data

This training focuses on the possibilities of utilizing large language models (LLMs) for interacting with private data. It introduces tools that harness the generative power of artificial intelligence in scenarios where data cannot leave the private environment at any point. The training covers the architecture and data requirements for creating a private ChatGPT, leveraging an understanding of the semantics of defined problems while maintaining control over private data.

The training includes the following topics:
- Overview of Large Language Models (LLMs)
- Introduction to LLMs and their capabilities in text generation and natural language processing.
- Transformers
- Discussion of transformer architecture, which forms the basis for LLMs.
- Presentation of the attention mechanism as a key element in LLMs, explaining how it allows models to selectively focus on specific information.
- Examples of attention model usage in LLMs.
- Various Methods for Customizing LLMs
- Introduction to different techniques used to customize LLMs for specific tasks, domains, and data.
- LLM Fine-Tuning
- The process of further training pre-trained LLMs on specific data.
- Utilizing fine-tuning to adapt the model for specific applications (practical exercises).
- Retrieval Augmented Generation (RAG)
- Explanation of the RAG approach, which combines text generation with information retrieval from an internal document set.
- Practical examples of RAG usage.
- Examples of advanced techniques to enhance RAG performance (practical exercises).

The training will be conducted in Polish or English, depending on the possible participation of non-Polish-speaking participants.
Duration: 6h (4h part 1 + 2h part 2)
First part: Wednesday, April 2, afternoon (15:00-19:00)
Second part: Thursday, April 3, (11:20-13:20)
Trainers:
- Rafał Możdżonek, a nuclear physicist by training, has worked for many years in data analysis, machine learning, and writing applications for scientific applications used at the National Center for Nuclear Research.
- Dr Krzysztof Nawrocki, High-energy particle physicist at the National Center for Nuclear Research with extensive experience in big data, grid computing, astronomy, and machine learning projects. Specialized in text data analysis, particularly within the medical field.

The number of participants is limited, priority of enrollment decides.

Presenters: Dr NAWROCKI, Krzysztof (NCBJ); MOŻDŻONEK, Rafał (NCBJ)

Session Classification: Wykorzystanie dużych modeli językowych dla danych prywatnych cz. 1 - LLM for private data, part 1