



Comtegra EDU

LLM RAG I FINETUNING

Zaprezentujemy Ci kompleksową wiedzę i wyposażymy Cię w umiejętności do samodzielnego projektowania, implementowania i optymalizacji rozwiązań opartych na dużych modelach językowych.

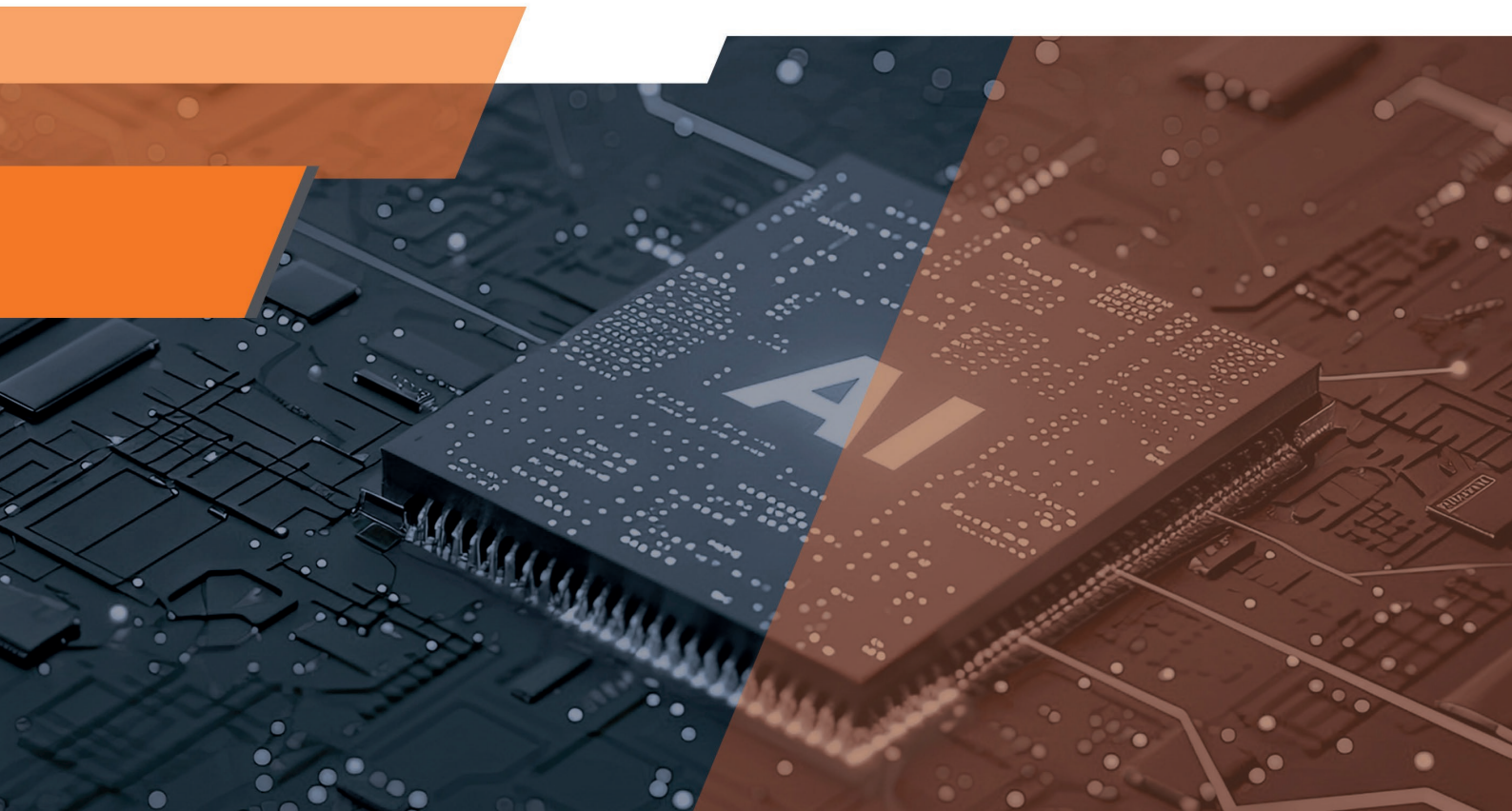
Zgłoś swój udział



szkolenia@comtegra.pl



Comtegra S.A. ul. Puławska 474, Warszawa





Dla kogo

To szkolenie będzie odpowiednie dla osób zainteresowanych technikami przetwarzania języka naturalnego, szczególnie dla programistów i inżynierów AI chcących zgłębić temat dużych modeli językowych (LLM).



Korzyści

Uczestnicy zyskają dogłębne zrozumienie funkcjonowania dużych modeli językowych i praktyczne umiejętności ich wykorzystania. Nauczą się tworzyć efektywne prompty, przeprowadzać fine-tuning modeli i implementować zaawansowane techniki, takie jak Retrieval Augmented Generation. Szkolenie wyposaży ich w narzędzia do budowy własnych aplikacji opartych na AI, w tym chatbotów i systemów automatycznego odpowiadania. Zdobyta wiedza obejmie również pracę z bazami wektorowymi, analizę danych, optymalizację wydajności modeli oraz przygotowanie danych zarówno dla rozwiązań RAG jak i finetuningu.



Język

Szkolenie prowadzone jest w języku polskim lub angielskim (w zależności od preferencji uczestników).



Forma

Warsztat składa się z 4 modułów. #1 jest wstępem teoretycznym - prezentacja. Kolejne moduły skonstruowane jako interaktywne lekcje z możliwością wykonywania kodu. Sesje hands on prowadzone są jako noteboki możliwe do uruchomienia za pomocą Jupyter Notebooka, VS Code i wielu innych. Po zakończonym warsztacie uczestnicy otrzymują wszystkie używane podczas szkolenia materiały do dalszej eksploracji.



1 Wstęp teoretyczny

- Prezentacja teoretyczna
 - Możliwości i ograniczenia modeli
 - Duże modele językowe
 - Przegląd modeli
 - Ranking LLMów - jaki model wybrać
 - Mechanizm uwagi
 - Infrastruktura
 - Orkiestracja
 - Bazy wektorowe
 - Retrieval Augmented Generation
- Wstęp do projektu

2 Prompty i dane

- Jak obsługiwać hostowane modele językowe
- Parametry odpowiedzi
- Zorganizuj swój kod - funkcje pomocnicze
- Loadery danych
- Czym są metadane
- Autoodpisywacz - projekt numer 1
- One-shot vs Few-shots

3 Finetuning

- Koncept matematyczny
- Pierwszy przykładowy finetuning
- Przygotowanie własnych danych pod finetuning
- Trening LoRA na prywatnym zbiorze danych
- Jak uruchomić swój model?
- Ewaluacja treningu
- Wznowienie treningu

4 Retrieval

- Wektorowe bazy danych
- Filtrowanie retriever-a
- Budowanie aplikacji RAG - projekt numer 2
- Praca z tekstem - jak przygotować dane
- Zaawansowane techniki wyciągania obiektów
- Modele Multimodalne