


Metryki

Last edited by **Paweł Sienkiewicz** 2 weeks ago

 [metryki_LLM.pdf](#)







Tytuł	Opis
 EVALUATING BERT AND GPT-2 MODELS FOR PERSONALISED LINKEDIN POST RECOMMENDATION	Artykuł ocenia skuteczność modeli BERT i GPT-2 w kontekście personalizowanych rekomendacji postów na LinkedIn. Badanie koncentruje się na analizie zdolności tych modeli do przetwarzania i zrozumienia kontekstowego treści, co jest kluczowe dla dostarczania trafnych rekomendacji użytkownikom. Wykorzystanie tych zaawansowanych modeli NLP ma na celu poprawę zaangażowania użytkowników poprzez dostarczanie im bardziej odpowiednich i spersonalizowanych treści.

LLM

Last edited by **Paweł Sienkiewicz** 3 weeks ago




<https://llama.meta.com/llama3/>

[LLama 3] (<https://medium.com/@akshithakumbam1/interesting-language-models-focus-on-llama-family-of-models-f7071a407a9c>)

Tytuł	Opis
 A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges	Wyjaśnia podstawowe koncepcje LLM, szczegółowo opisując tradycyjne fazy treningu i istotną ewolucję ich architektur, szczególnie skupiając się na modelach "transformers". Omawia różnorodne zasoby i zestawy danych używane w treningu LLM, oraz porównuje różne modele do siebie. Różne metody treningowe, techniki optymalizacji i zastosowania są również omówione.
 Attention is all you need	Praca która jako pierwsza przedstawiła model "transformers". Dokładnie omawia architekturę tego modelu, oraz porównuje jego wyniki do wcześniej używanych architektur.
 A Neural Probabilistic Language Model	Praca przedstawiająca pierwszy model językowy oparty na sieciach neuronowych. Model uczy się jednocześnie rozproszonych reprezentacji słów i funkcji prawdopodobieństwa dla sekwencji słów, co poprawia generalizację, zwłaszcza w przypadku nowych, nieznanych sekwencji.
 TinyStories: How Small Can Language Models Be and Still Speak Coherent English?	Bada możliwości małych modeli językowych w generowaniu spójnych i zrozumiałych tekstów. Autorzy stworzyli syntetyczny zestaw danych, zawierający krótkie opowiadania napisane prostymi słowami zrozumiałymi dla dzieci w wieku 3-4 lat, wygenerowany przez GPT-3.5 i GPT-4. Wykazano, że modele językowe poniżej 10 milionów parametrów lub z prostymi architekturami są w stanie tworzyć płynne, spójne i gramatycznie poprawne teksty, wykazując również pewne zdolności rozumowania
 Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone	Przedstawia model językowy Phi-3-mini, który ma 3,8 miliarda parametrów i został przeszkolony na 3,3 bilionach tokenów. Model ten osiąga wydajność porównywalną z większymi modelami, takimi jak GPT-3.5, jednocześnie będąc na tyle kompaktowym, że może działać na telefonie komórkowym. Model wykorzystuje zaawansowane techniki, takie jak lekkie mechanizmy uwagi, destylacja wiedzy i kwantyzacja.
 Zwiększenie context window modelu llama 3 z 8192 tokenów do 80 tysięcy	Opisuje metodę zwiększenia długości kontekstu modelu Llama-3-8B-Instruct z 8K do 80K za pomocą techniki fine-tuning QLoRA. Proces treningowy jest bardzo efektywny i osiąga znakomite wyniki w szerokim zakresie zadań związanych z długim kontekstem, jednocześnie zachowując zdolności do pracy z krótszym kontekstem.

Fine Tuning






Last edited by **Paweł Sienkiewicz** 3 weeks ago

Tytuł	Opis
 Fine tuning przy pomocy metody EMA-TARGET SELF-SUPERVISED LOSS	Bada metodę wykorzystania średniej wykładniczej (EMA) jako celu do samonadzorowanego uczenia się w celu ulepszenia fine-tuningu wcześniej przeszkolonych modeli mowy. Podejście to ma na celu lepsze zachowanie pierwotnie nauczonych reprezentacji podczas dostosowywania modelu do konkretnych zadań.
 Designing a Fine-Tuning Tool for Machine Learning with High-Speed and Low-Power Processing	koncentruje się na opracowaniu narzędzia do fine-tuningu modeli uczenia maszynowego, które charakteryzuje się wysoką szybkością działania i niskim zużyciem energii. Główne innowacje obejmują wykorzystanie architektury ASIC-FPGA oraz technik faloprzewodowych, co pozwala na szybsze przetwarzanie danych przy jednoczesnym zmniejszeniu zużycia energii.
 A fine-tuning approach research of pre-trained model with two stage	Przedstawia dwustopniowe podejście do fine-tuningu, obejmujące wstępne dostrajanie końcowej warstwy z użyciem straty wagowej, oraz standardowe dostrajanie całego modelu w drugim etapie. Poprawa wydajności modeli w radzeniu sobie z niezbalansowanymi danymi i lepsze wykorzystanie reprezentacji wstępnie wytrenowanych.

Educational Materials

Last edited by **Paweł Sienkiewicz** 2 weeks ago


 [Sprawozdanie_z_wirtualnych_warsztatów_CLARIN-PL_w_praktyce_badawczej_19.20.11.2020_Wrocław_Politechnika_Wrocławska.pdf](#)

Tytuł	Opis
 A Deep Learning based Self-Assessment Tool for Personality Traits and Interview Preparations	Przedstawia narzędzie oparte na głębokim uczeniu, które pomaga użytkownikom ocenić swoje cechy osobowości oraz przygotować się do różnych rodzajów rozmów kwalifikacyjnych.
 Resolving Single-Sentence Answers for the development of an Automated Assessment Process	Przedstawia metodę automatycznej oceny krótkich odpowiedzi pisemnych. Celem jest stworzenie systemu, który potrafi dokładnie ocenić odpowiedzi na pytania w formie jednozdaniowej, co jest kluczowe w kontekście edukacyjnym i egzaminacyjnym. W pracy opisano zastosowanie zaawansowanych algorytmów przetwarzania języka naturalnego i technik głębokiego uczenia do analizy i oceny takich odpowiedzi, co pozwala na szybkie i dokładne przeprowadzanie egzaminów online.
 sztuczna inteligencja w dydaktyce	Zbiór prac naukowych np. "Sztuczna inteligencja i jej wykorzystanie w logistyce ." lub "Możliwości wykorzystania sztucznej inteligencji w działalności dydaktycznej i naukowej"
 Resume Building Application based on LLM (Large Language Model).	Praca koncentruje się na automatyzacji i personalizacji procesu tworzenia profesjonalnych życiorysów. Aplikacja wykorzystuje zaawansowane algorytmy NLP do analizy i interpretacji informacji dostarczanych przez użytkownika, generując spersonalizowane, dobrze sformułowane CV dostosowane do konkretnej branży lub stanowiska. System analizuje doświadczenie zawodowe, umiejętności i osiągnięcia użytkownika, aby stworzyć optymalny dokument, który wyróżnia się na rynku pracy, jednocześnie oszczędzając czas i wysiłek użytkownika.
 GPT-3 Powered System for Content Generation and Transformation	Opisuje system oparty na GPT-3 do generowania i transformacji treści wykorzystuje zaawansowane możliwości modelu GPT-3 do automatycznego tworzenia, modyfikowania i ulepszania tekstów.
https://openai.com/index/summarizing-books/	Sposób na streszczanie książek/długich tekstów. Ponieważ niektóre teksty są za długie, aby context window pomieściło wszystkie informacje potrzeba odpowiednich sposobów na streszczanie tych tekstów.
https://platform.openai.com/docs/guides/embeddings/what-are-embeddings	O embeddingach

Transformer


Last edited by **Paweł Sienkiewicz** 2 weeks ago

https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

Tytuł	Opis
 Attention Is All You Need	Praca która jako pierwsza przedstawiła model "transformers". Dokładnie omawia architekturę tego modelu, oraz porównuje jego wyniki do wcześniej używanych architektur.



Przetwarzanie języka naturalnego

Last edited by Paweł Sienkiewicz 2 weeks ago

Tytuł	Opis
 O wiele więcej niż Google Translate, czyli komputerowe przetwarzanie języka naturalnego (NLP) w translatoryce i translatologii	<p>Opisuje ewolucję narzędzi NLP, które początkowo koncentrowały się na podstawowych zadaniach tłumaczeniowych, a teraz oferują zaawansowane funkcje analizy językowej i semantycznej.</p> <p>Podkreśla zastosowanie NLP w badaniach nad tłumaczeniem, w tym w analizie jakości tłumaczeń, automatycznym wykrywaniu błędów tłumaczeniowych oraz w tworzeniu korpusów porównawczych dla badaczy. Analizuje aktualne wyzwania stojące przed technologią NLP, takie jak wielojęzyczność, zrozumienie kontekstu kulturowego, a także etyczne aspekty użycia NLP w tłumaczeniach.</p>

Prompt Engineering

Last edited by Paweł Sienkiewicz 4 days ago

Tytuł	Opis
 Chain-of-Thought Prompting Elicits Reasoning in Large Language Models	Artykuł bada technikę "Chain-of-Thought Prompting" (CoT), która ma na celu wywoływanie zdolności do rozumowania w dużych modelach językowych (LLMs), takich jak GPT-3. Technika ta polega na dostarczaniu modelowi serii powiązanych wskazówek, które prowadzą go przez proces rozumowania, pozwalając na bardziej skomplikowane i logiczne odpowiedzi. CoT wykazuje potencjał w poprawie wydajności LLMs w zadaniach wymagających myślenia wieloetapowego, zwiększając ich zdolność do rozwiązywania problemów i podejmowania decyzji.
https://platform.openai.com/docs/guides/prompt-engineering/strategy-test-changes-systematically	Strona na której jest bardzo dużo prac naukowych
 Using ChatGPT Standard Prompt Engineering Techniques in Lesson Preparation: Role, Instructions and Seed-Word Prompts	Praca opisuje techniki kreowania instrukcji do przygotowania lekcji.