Veľký jazykový model (Large Language Model, LLM), self supervised, tooltip, transformer architektúra, pre-training, fine-tuning, inference, few-shot learning, emergentné správanie :)

Veľký jazykový model v oblasti umelej inteligencie je jazykový model trénovaný na obrovskom množstve textu pomocou špeciálneho strojového učenia bez učiteľa (self-supervised)[1].

Ide o generatívny model, ktorý predikuje nasledujúce slovo alebo vetu na základe kontextu.

Architektúra a technológia

Je založený na **transformer architektúre[2]. Využíva mechanizmus self-attention[3] na analýzu kontextu a paralelizáciu výpočtov, čo umožňuje pracovať s dlhými vstupmi efektívnejšie než staršie modely (napríklad RNN).**

Obsahuje **miliardy až bilióny parametrov**, vďaka čomu dokáže pochopiť gramatiku, význam aj jemné jazykové kontexty.

Ako LLM funguje?

1. fáza resp. Predtréning (pre-training) - model sa učí predikovať ďalšie slovo v sekvencii z rozsiahlych textových datasetov (knihy, weby, články...).
2. fáza resp. Doladenie (fine-tuning) - následne sa model môže dolaďovať na konkrétne úlohy. Napríklad na preklad, sumarizáciu, generovanie kódu, otázky a odpovede.
3. fáza resp. Interferencia[4] (inference) - pri použití generuje text, token za tokenom, na základe vstupného promptu, predikujúc pravdepodobne najvhodnejšie pokračovanie.

Schopnosti a aplikácie:

- **generovanie textu**: písanie esejí, článkov, kódu, poézie,
- **preklad a sumarizácia**: automatický preklad jazykov a tvorba skrátených verzií textu,
- **otázky a odpovede**: poskytovanie informatívnych odpovedí na ľudské otázky,
- **rôzne úlohy**: analýza sentimentu[5], tvorba tooltipov[6], právne klauzuly, technická dokumentácia...

Význam a vývoj:

- predstavujú **prelom v NLP**, umožnili vznik chatbotov ako ChatGPT, Gemini, Claude či Copilot, schopných všeobecných textových úloh bez špeciálneho programovania pre každú problematiku,
- modely sú schopné **few-shot learning[7] či emergentného správania[8] pri dostatočne veľkom trénovacom datasete, tzv. mega dátach a veľkej architektúre,**
- **LLM znamenajú nový štandard v spracovaní prirodzeného jazyka**, umožňujúci široké využitie v edukačných, podnikových aj kreatívnych aplikáciách.

[1] Self-supervised je špeciálny typ učenia bez učiteľa, ktorý je veľmi populárny pri trénovaní LLM, vizuálnych modelov a multimodálnych systémov.
[2] Transformer architektúra je základná štruktúra moderných jazykových modelov; ako LLM; navrhnutá na spracovanie sekvencií (postupností, napríklad textu) tak, aby pracovala efektívne a paralelne.
[3] Mechanizmus v architektúre transformerov, ktorý umožňuje modelu „pozerať sa“ na všetky časti vstupnej sekvencie a určiť, ktoré slová (tokeny) sú pre aktuálny výpočet najdôležitejšie.
[4] Inferencia je časť procesu umelej inteligencie (LLM), keď už je model natrénovaný a používa sa na predpovedanie alebo generovanie výstupov na základe vstupných dát. V kontraste s tréningovou fázou, kde sa model učí z dát, v inferenčnej fáze model neučí, ale aplikuje naučené vzory.
[5] Analýza sentimentu je technika spracovania prirodzeného jazyka (NLP), ktorá zisťuje emocionálny tón alebo postoj v texte. Cieľom je určiť, či je obsah pozitívny, negatívny alebo neutrálny, prípadne jemnejšie kategórie. Napríklad radosť, hnev, smútok.
[6] Tooltip je malý informačný prvok v používateľskom rozhraní, ktorý sa zobrazí, keď používateľ prejde kurzorom nad prvok. Napríklad ikonu, tlačidlo alebo text.

Copyright © 2013 - 2026 Encyklopédia poznania

[7] Few-shot learning je spôsob, ako veľké jazykové modely (LLM) dokážu riešiť nové úlohy s veľmi malým počtom príkladov (tzv. „shots“), ktoré im dáte priamo v promptoch.
[8] Emergentné správanie v kontexte veľkých jazykových modelov (LLM) znamená, že model vykazuje schopnosti alebo vlastnosti, ktoré neboli explicitne naprogramované ani očakávané počas tréningu, ale „objavia sa“ pri dostatočne veľkej škále modelu alebo dát.

Copyright © 2013 - 2026 Encyklopédia poznania