

**KATTA TIL MODELLARI (LLM): TASNIFI, AFZALLIKLARI,
CHEKLOVLARI VA QO‘LLANILISH SOHALARI**

Matyakubova Noila Shakirjanovna,
Alfraganus University, Tillarni o‘qitish markazi

Annatotsiya

Ushbu maqolada katta til modellari (Large Language Models – LLM)ning ilmiy-nazariy tasnifi, ularning ustuvor jihatlari, mavjud cheklovlari hamda hozirgi davrda qo‘llanilayotgan asosiy amaliy yo‘nalishlari tahlil qilinadi. LLMlar sun‘iy intellektning chuqur o‘rganishga asoslangan eng muhim texnologiyasi bo‘lib, ular tilning sintaktik, semantik va pragmatik qatlamlarini o‘rganish orqali ko‘plab murakkab vazifalarni amalga oshirishga qodir. Shunga qaramay, LLMlar ma‘lumotga qaramlik, hisoblash xarajatlari va etik masalalar tufayli bir qator cheklovlarga ham ega. Maqolada ushbu texnologiyaning ilmiy asoslarini yoritish bilan birga, uning imkoniyatlari va cheklovlarini ham tizimli tarzda ko‘rib chiqiladi.

Kalit so‘zlar: Large Language Model, NLP, transformer arxitekturası, generativ AI, til modellari, ustuvorliklar, cheklovlar.

Kirish

Sun‘iy intellektning jadal rivojlanishi natijasida katta til modellari (LLM) zamonaviy raqamli texnologiyalar markaziga aylandi. Bu modellar o‘ta katta hajmdagi matnli ma‘lumotlarga tayangan holda tilning ichki statistik qonuniyatlarini o‘rganadi va shu asosda inson nutqini qayta yaratish, tushunish va modellashtirishga qodir bo‘ladi. LLM texnologiyasi ChatGPT, GPT-4, LLaMA, BERT, T5 kabi mashhur modellarda o‘z ifodasini topgan.

Katta til modellari ya‘ni LLM (Large Language Models) – bu milliardlab parametrga ega bo‘lgan, sun‘iy intellektning chuqur o‘rganishga asoslangan til modellari bo‘lib, ularga juda katta hajmdagi matnlar (internet, kitoblar, maqolalar, dialoglar va h.k.) orqali o‘rgatiladi. LLMlarning asosiy vazifasi tilning ichki qonuniyatlarini, ma‘no munosabatlarini, kontekstni va nutq mantiqini o‘rganish.

LLMlarning matn generatsiyasi, savol-javob, tarjima, muloqot, kod yozish, xulosalash kabi funksiyalari ularni universal intellektual tizim sifatida talqin qilishga asos bo‘ladi. Shu bilan birga, bunday texnologiyalarni to‘g‘ri tasniflash,

afzallik va cheklovlarini aniqlash ilmiy va amaliy jihatdan dolzarb masala bo'lib qolmoqda.

LLMlar asosan katta hajmdagi internet matnlari, ochiq korpuslar va yirik annotatsiyalangan ma'lumotlar bazalari asosida o'qitiladi. Bu ma'lumotlar ko'proq xalqaro kommunikatsiyada ustun mavqega ega bo'lgan tillarga tegishli bo'lgani uchun, kam resursli tillarda ishlashda modelning ishlashi sezilarli darajada pasayadi, turli xil sistematik xatolar, biaslar va madaniy nomuvofiqliklar yuzaga keladi.

LLM tushunchasi va rivojlanish bosqichlari

Katta til modellari – bu milliardlab parametrli, chuqur neyron tarmoqlarga asoslangan statistik modellar bo'lib, katta hajmdagi matnlar ustida oldindan o'qitiladi. Ular tilning ehtimollik taqsimotini o'rganish orqali kontekstga mos, grammatik jihatdan to'g'ri va ma'no jihatidan bir butun bo'lgan matnlar yaratishga qodir. LLMlar GPT, BERT, LLaMA, Mistral, FLAN-T5, BLOOMZ kabi modellarga tayanadi va hozirgi NLPning ko'plab tizimlarining markazida turadi. LLMlarning shakllanishida uch asosiy bosqichni ko'rsatish mumkin:

- **Oldindan o'qitilgan til modellari (pre-trained language models)** – BERT, GPT-2 kabi modellar umumiy korpuslarda o'qitilib, keyinchalik aniq vazifaga moslab qayta o'qitiladi (***fine-tuning***).
- **Ko'p tilli va ko'p domenli modellar** – mBERT, XLM-R, mT5 kabi modellar ko'plab tillarni bir vaqtda qamrab olib, ***cross-lingual transfer*** imkonini yaratadi.
- **Inson fikriga moslashtirilgan generativ LLMlar** – ChatGPT, GPT-4 turkumidagi modellar ***Reinforcement Learning from Human Feedback (RLHF)*** kabi metodlar yordamida muloqot, dialog, interaktiv vazifalarga moslashtirildi.

LLMlarning NLPdagi ahamiyati

LLMlar NLP texnologiyalarining samaradorligi va qo'llanish doirasini sezilarli darajada kengaytirdi. Ushbu modellarning paydo bo'lishi matn bilan ishlashning an'anaviy usullarini yangi bosqichga olib chiqdi va NLP sohasida bir nechta tub burilishlarni yuzaga keltirdi. U universal platforma sifatida qo'llanilib bitta katta model asosida bir vaqtning o'zida bir nechta vazifalarni ya'ni tarjima, xulosa, savol-javob, sentiment tahlili va boshqalarni bajarish mumkin bo'ldi. Bu

yondashuv ananaviy “har vazifa uchun alohida model” paradigmasidan keskin farq qiladi.

Bundan tashqari annotatsiyasiz o‘rganish imkoniyatini taqdim etadi. Katta hajmdagi qo‘lda belgilangan (label) ma’lumot talabini sezilarli kamaytirdi. Zero-shot va few-shot o‘qitish yordamida kam misol bilan ham yangi vazifaga moslashish imkoni paydo bo‘ldi. Transfer va cross-lingual imkoniyatlar paydo bo‘ldi. Ko‘p tilli o‘qitish natijasida bir tilda olingan bilimlar boshqa tilda ham qisman ishlay boshladi, bu esa kam resursli tillar uchun juda muhim hisoblanadi. Interaktiv va adaptiv tizimlar yaratildi. Chatbotlar, intellektual yordamchilar, ta’lim va sog‘liqni saqlash tizimlari LLMlar asosida yanada moslashuvchan va kontekstga sezgir shaklda rivojlanmoqda.

LLMlarning muhim xususiyati shundaki, ular oldindan qat’iy qoidalarsiz, kontekstga moslashgan holda matnni tahlil qilish va yaratish imkoniyatiga ega. Bu esa ularni klassik NLP yondashuvlaridan keskin farqlaydi. LLMlar NLP sohasidagi ko‘plab asosiy vazifalarni yuqori aniqlik bilan bajarishga imkon bermoqda. Jumladan:

- matnni avtomatik tarjima qilish;
- savol-javob tizimlarini yaratish;
- matnni umumlashtirish (summarizatsiya);
- sentiment tahlili;
- nutqni tushunish va mantiqiy xulosalar chiqarish.

Avvallari har bir vazifa uchun alohida model ishlab chiqish talab etilgan bo‘lsa, LLMlar ko‘p vazifali (multi-task) yondashuv orqali bir model doirasida turli NLP masalalarini hal qilish imkonini bermoqda.

LLMlarning NLPdagi eng katta afzalliklaridan biri bu kontekstual tushunish qobiliyatidir. Ular so‘zlarni alohida emas, balki gap va matn doirasidagi o‘zaro bog‘liqlikda tahlil qiladi. Natijada so‘zning ma’nosi kontekstga qarab to‘g‘ri aniqlanadi. Bu esa sinonimlar, ko‘p ma’noli so‘zlar va murakkab jumlar bilan ishlashda yuqori aniqlikni ta’minlaydi.

Bugungi kunda LLMlar turli sohalarda keng qo‘llanilmoqda. Masalan, ta’lim tizimida aqlli yordamchilar va o‘quv platformalarida, tibbiyotda hujjatlarni tahlil qilish va klinik qarorlarni qo‘llab-quvvatlashda, biznesda mijozlar bilan muloqot qiluvchi chatbotlar orqali, ilmiy tadqiqotlarda katta hajmdagi matnli ma’lumotlarni tahlil qilishda. Bu holat LLMlarning nafaqat nazariy, balki amaliy jihatdan ham yuqori ahamiyatga ega ekanligini ko‘rsatadi.

Kam resursli tillar uchun LLMlardan foydalanishning muammolari va ijtimoiy masalalar

Sun'iy intellekt va tabiiy tilni qayta ishlash (NLP) sohasida katta til modellari muhim yutuqlarga erishgan bo'lsa-da, ularning samaradorligi barcha tillar uchun bir xil darajada emas. Ayniqsa, kam resursli tillar uchun LLMlardan foydalanishda texnik, lingvistik va ijtimoiy muammolar yuzaga kelmoqda. Ushbu masalalar raqamli tengsizlikni kuchaytirishi va ayrim tillarning texnologik rivojlanishdan chetda qolishiga olib kelishi mumkin.

Kam resursli tillar – bu raqamli muhitda yetarli hajmdagi matn korpuslari, belgilangan (annotatsiyalangan) ma'lumotlar, lug'atlar va til texnologiyalari mavjud bo'lmagan tillardir. Bunday tillar ko'pincha mahalliy yoki mintaqaviy miqyosda qo'llaniladi va global internet kontentida kam uchraydi. Natijada LLMlarni o'qitish jarayonida ushbu tillar yetarlicha qamrab olinmaydi.

Ko'plab kam resursli tillarida raqamli matnlar, standart imlo, normativ korpuslar kabi ma'lumot yetishmasligi muammosi mavjud. LLMlar asosan katta hajmdagi ma'lumotlarga tayangan holda o'qitiladi. Kam resursli tillar uchun sifatli va xilma-xil matnlar yetishmasligi modelning tilni to'liq o'rganishiga to'sqinlik qiladi. Bu esa grammatik xatolar, semantik noto'g'ri talqinlar va kontekstni yetarli darajada tushunmaslik kabi muammolarni keltirib chiqaradi. Bundan tashqari, kam resursli tillarda dialektlar va yozuv variantlarining ko'pligi ma'lumotlarni standartlashtirishni qiyinlashtiradi. Natijada LLMlar tilning rasmiy shakliga nisbatan haddan tashqari moyil bo'lib, og'zaki yoki mintaqaviy variantlarni e'tibordan chetda qoldirishi mumkin.

Til – bu faqat muloqot vositasi emas, balki madaniyat va milliy identitetning muhim tarkibiy qismidir. Agar LLMlar kam resursli tillarni yetarli darajada qo'llab-quvvatlamasa, bu tillarning raqamli makondagi mavqei pasayadi. Natijada yosh avlod orasida ona tilidan foydalanish qisqarishi va tilning sekin-asta texnologik muhitdan siqib chiqarilishi ehtimoli ortadi. Bundan tashqari, madaniy kontekstni to'liq tushunmaslik LLMlar tomonidan yaratilgan matnlarning mahalliy qadriyatlar va an'analarga mos kelmasligiga olib kelishi mumkin. Shuningdek, bias va stereotiplar ya'ni genderga oid stereotiplar, mintaqaviy nomuvofiqliklar, ijtimoiy guruhlariga nisbatan adolatsiz tasvirlar ko'plab ishlarda qayd etilgan. Xavfsizlik muammolari, modelni manipulyatsiya qilib, istalmagan natijalar olish mumkinligi alohida tahdid sifatida ko'rsatilgan.

Kam resursli tillar uchun LLMlardan foydalanishda axloqiy masalalar ham muhim o'rin tutadi. Ma'lumotlarni yig'ish jarayonida mualliflik huquqi, shaxsiy

ma'lumotlar himoyasi va rozilik masalalari yetarlicha nazorat qilinmasligi mumkin. Ayniqsa, kichik til jamoalarida bunday muammolar jiddiy ijtimoiy oqibatlariga olib kelishi ehtimoli mavjud. Shu bilan bir qatorda, texnologik infratuzilmaning cheklanganligi sababli, kam resursli tillarda so'zlashuvchi jamiyatlar LLMlarga asoslangan xizmatlardan teng darajada foydalana olmasligi mumkin. LLMlar katta hisoblash quvvatini, xotira va energiya talab qiladi; resurslari cheklangan mintaqalarda bu ham alohida muammo hisoblanadi. Shu sababli kam resursli tillarda LLMlardan foydalanishda etik me'yorlar, ijtimoiy adolat, til siyosati, mahalliy hamjamiyat bilan hamkorlik kabi jihatlarni hisobga olish zarur.

Xulosa

Xulosa qilib aytganda, katta til modellari NLP sohasining rivojlanishida muhim strategik o'rin egallaydi. Ular tilni chuqur tushunish, moslashuvchanlik va ko'p funksiyalilik xususiyatlari orqali tabiiy til bilan ishlash samaradorligini yangi bosqichga olib chiqdi. Kelajakda LLMlar yanada takomillashib, inson va kompyuter o'rtasidagi muloqotni yanada tabiiy va samarali qilishiga shubha yo'q. Ammo, LLMlar kam resursli tillar uchun katta imkoniyatlar bilan birga jiddiy muammolarni ham yuzaga keltiradi. Ma'lumotlar yetishmasligi, til notengligi, madaniy va axloqiy masalalar ushbu texnologiyalarni adolatli va samarali joriy etishda asosiy to'siqlar hisoblanadi. Shu sababli, kelajakda kam resursli tillar uchun maxsus korpuslar yaratish, mahalliy mutaxassislarni jalb qilish va ijtimoiy-madaniy omillarni hisobga olgan holda LLMlarni rivojlantirish muhim ahamiyat kasb etadi.

Foydalanilgan adabiyotlar ro'yhati

1. Kaplan, J., McCandlish, S., Henighan, T., et al. (2020). Scaling Laws for Neural Language Models. arXiv:2001.08361.
2. Hadi, Muhammad Usman & Al-Tashi, Qasem & Qureshi, Rizwan & Shah, Abbas & Muneer, Amgad & Irfan, Muhammad & Zafar, Anas & Shaikh, Muhammad & Akhtar, Naveed & Wu, Jia & Mirjalili, Seyedali. (2023). Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. 10.36227/techrxiv.23589741.
3. Hu, E. J., Shen, Y., Wallis, P., et al. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685.

4. Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training Language Models to Follow Instructions with Human Feedback (InstructGPT/RLHF). NeurIPS.
5. Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022). Training Compute-Optimal Large Language Models (Chinchilla). NeurIPS.
6. Touvron, H., Lavril, T., Izacard, G., et al. (2023). LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.13971.
7. Zhao, W. X., Zhou, K., Li, J., et al. (2023). A Survey of Large Language Models. arXiv:2303.18223.
8. Huang, L., Yu, W., Ma, W., et al. (2023). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. arXiv:2311.05232.